

Genomics Review of Holocellulose Deconstruction by Aspergilli

Fernando Segato,^{a*} André R. L. Damásio,^{a,b} Rosymar C. de Lucas,^{a*} Fabio M. Squina,^b Rolf A. Prade^{a,b}

Department of Microbiology & Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA^a; Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil^b

SUMMARY	588
INTRODUCTION	589
HOLOCCELLULOSE ENZYME BREAKDOWN SYSTEMS	589
ENZYME STRUCTURE-FUNCTION AND SUBSTRATE RELATIONSHIPS	591
Cellulases	593
Cellobiohydrolases	593
Endoglucanases	595
Copper-dependent lytic polysaccharide monooxygenases	595
Celllobiose dehydrogenase	596
β -Glucosidases	598
Aspergillus cellulose degradation summary	598
Hemicellulases	599
Hemicellulose backbone-hydrolyzing enzymes	600
(i) Endo-1,4- β -xylanases	600
(ii) β -Xylosidases (4- β -D-xylan xylohydrolases)	600
(iii) Mannanases and β -mannosidase	600
Side chain-hydrolyzing enzymes	600
(i) Arabinoxyranosidases	600
(ii) Arabinanases	603
(iii) β -Glucuronidases and xylan α -1,2-glucuronidase	603
(iv) Xyloglucanases	603
Accessory enzymes	604
(i) Acetyl xylan esterases	604
(ii) Ferulic acid esterases	604
Aspergillus hemicellulase degradation summary	604
CONCLUSIONS	606
ACKNOWLEDGMENTS	606
REFERENCES	606
AUTHOR BIOS	613

SUMMARY

Biomass is constructed of dense recalcitrant polymeric materials: proteins, lignin, and holocellulose, a fraction constituting fibrous cellulose wrapped in hemicellulose-pectin. Bacteria and fungi are abundant in soil and forest floors, actively recycling biomass mainly by extracting sugars from holocellulose degradation. Here we review the genome-wide contents of seven *Aspergillus* species and unravel hundreds of gene models encoding holocellulose-degrading enzymes. Numerous apparent gene duplications followed functional evolution, grouping similar genes into smaller coherent functional families according to specialized structural features, domain organization, biochemical activity, and genus genome distribution. Aspergilli contain about 37 cellulase gene models, clustered in two mechanistic categories: 27 hydrolyze and 10 oxidize glycosidic bonds. Within the oxidative enzymes, we found two cellobiose dehydrogenases that produce oxygen radicals utilized by eight lytic polysaccharide monooxygenases that oxidize glycosidic linkages, breaking crystalline cellulose chains and making them accessible to hydrolytic enzymes. Among the hydrolases, six cellobiohydrolases with a tunnel-like structural fold embrace single crystalline cellulose chains and cooperate at nonreducing or reducing end termini, splitting off cellobiose. Five endoglucanases group into four structural families

and interact randomly and internally with cellulose through an open cleft catalytic domain, and finally, seven extracellular β -glucosidases cleave cellobiose and related oligomers into glucose. Aspergilli contain, on average, 30 hemicellulase and 7 accessory gene models, distributed among 9 distinct functional categories: the backbone-attacking enzymes xylanase, mannosidase, arabinanase, and xyloglucanase, the short-side-chain-removing enzymes xylan α -1,2-glucuronidase, arabinofuranosidase, and xylosidase, and the accessory enzymes acetyl xylan and feruloyl esterases.

Address correspondence to Rolf A. Prade, prade@okstate.edu.

* Present address: Fernando Segato, Departamento de Biotecnologia da Escola de Engenharia de Lorena, Universidade de São Paulo, Lorena, SP, Brazil; Rosymar C. de Lucas, Novozymes Brazil, Araucária, PR, Brazil.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/MMBR.00019-14>.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

[doi:10.1128/MMBR.00019-14](https://doi.org/10.1128/MMBR.00019-14)

INTRODUCTION

Plants capture photons from sunlight to fix carbon dioxide into sugars via the Calvin cycle, and they deposit most of the carbohydrates onto plant cell walls (1). Plant cell walls are major constituents of what is commonly defined as biomass, a huge component of all terrestrial habitats that is mostly represented by the cell walls of dead plants, trees, grasses, and leftovers of actively managed croplands. Biomass is also a natural waste product that accumulates along sugar mills and alcohol refineries worldwide (2, 3).

However, plants, with their robust and refractory cell walls, die and are completely recycled in nature. Polysaccharides are deconstructed and resulting sugars assimilated by all microorganisms that live in soil and forest floors (4). Furthermore, natural decomposition of woody (most of the available biomass) materials remains a recalcitrant process that takes place in soil microbiomes and perhaps employs numerous enzymes derived from not one organism but a whole community of microbes (5).

Bacteria are major components of soil environments and are equipped with many of the enzymes needed to degrade plant cell walls (6, 7). Fungi are also present in soil and in the rumina of herbivores and are rich in polysaccharide-degrading enzymes and active in recycling plant cell walls (8, 9). For example, *Aspergillus nidulans* grown on sorghum stover under solid-state culture conditions secreted a total of 294 proteins, predominantly hemicellulases, cellulases, polygalacturonases, chitinases, esterases, and lipases, over a 2-week period, while the fungus used only 30% of the total available hemicellulose and cellulose accessible from sorghum stover (10).

However, not all fungi are alike; some fungi, such as *Trichoderma reesei* and *Aspergillus niger*, employ predominantly a classical acid catalysis hydrolytic model of degrading plant cell wall polymers, while others, such as *Myceliophthora thermophila* (11, 12) and *Phanerochaete chrysosporium* (13, 14), appear to take advantage of an oxidative route of breaking down glycoside bonds. Even though some fungi predominantly hydrolyze biomass while others employ an oxidative mechanism, most fungi contain a mixed set of hydrolases and oxidases. *Neurospora crassa* (15, 16) and *Aspergillus* species appear to harbor a mixed acid hydrolysis and oxidative system.

In the last decade, biomass degradation by fungal enzymes has received renewed attention because of the need to utilize biomass as a food source to drive fermentation processes that manufacture biofuels and other feedstocks for chemicals and pharmaceuticals. Seamless integration of a biomass-to-sugar conversion process with traditional sugar-to-ethanol fermentation is the bottleneck of the biofuel industry (17–23). Biomass carbohydrate heterogeneity and structural fiber complexity (recalcitrance) result in water-deprived structures that hinder the access of enzymes.

The genomic DNA sequences of numerous *Aspergillus* species have been determined, including *Aspergillus oryzae* (24), *Aspergillus fumigatus* (25), and *Aspergillus niger* (26). Genome-wide reports comparing *Aspergillus flavus* with *A. oryzae* (27) and *A. nidulans* with *A. fumigatus* and *A. oryzae* (28) and a curated comparative genomics resource for aspergilli (29) are available as well. Several articles describing plant cell wall-degrading enzymes have been published, mainly focusing on glycoside hydrolases (30, 31) or fungal sets of plant polysaccharide-degrading enzymes (32, 33), as well as comparisons of *A. nidulans* with *A. niger* and *A. oryzae* (34–36). The last comprehensive review of cellulose- and

hemicellulose-degrading enzymes produced by *Aspergillus* was published in 2001 (37).

Here we propose a comprehensive genomic review of seven sequenced *Aspergillus* genomes and compare them with the model biomass degrader *T. reesei* and the genetic model system *N. crassa*. In examining gene models encoding cellulases and hemicellulases in fungi, one encounters gene duplications and protein products with overlapping biochemical functions, illustrating an evolutionary effort to acquire functions capable of degrading the recalcitrant plant cell wall. As a result, the combination of genetic multiplicity (gene duplications or multiple acquisitions of the same gene, followed by evolution into specialized functions) and functional redundancy (genes with diversified biochemical functions, such as exo- and endohydrolysis of the same substrate) creates a complex and overlapping repertoire of enzymes, for which we provide a comprehensive overview.

HOLOCCELLULOSE ENZYME BREAKDOWN SYSTEMS

Holocellulose is the carbohydrate fraction of biomass (lignocellulose) that includes the total polysaccharide fraction obtained after extractives and lignin have been removed from natural materials. (A rough schematic is shown in Fig. 1.) Holocellulose (~65 to 85%) contains polymers, such as cellulose (~30%), hemicellulose (~20 to 30%), and pectin (~5 to 30%), that incorporate sugars as their basic repeating units (38). These sugars represent a massive pool of C—H-bonded energy molecules locked into recalcitrant polymers.

Cellulose, hemicellulose, and lignin quantities are variable in plants (Table 1). For example, in sugar cane bagasse, the main hemicellulose is arabinoxylan, and in soft- and hardwoods, hemicellulose contains glucomannan and arabinan (39, 40). In addition, pectin also contains a carbohydrate-formulated polymeric complex. However, it appears at various rates in different plant tissues. While abundant in fruits, very little pectin is present in traditional biomass, such as wood, sorghum, and corn stover or sugar cane bagasse. Thus, the main pectin-degrading activities lie outside the scope of this review.

Cellulose, a linear polymer, consists of D-glucose units connected to each other by glycoside β -1,4 linkages. Each glucose unit is rotated 180° from its neighbor molecule, enabling long linear chains with lengths of 2,000 to 25,000 glucose residues. Typical cellulose microfibrils are composed of 36 parallel cellulose chains in a paracrystalline, linear, hexagonal arrangement with a 10-nm diameter, known as type I cellulose (41). The hydrogen bonds between adjacent cellulose polymers form crystalline structures that give plants structural strength. Cellulose occurs in helically wound reinforcing crystalline microfibrils as well as in an amorphous form glued with lignin and hemicellulose to serve as a support matrix (42, 43). Cellulosic microfibrils are approximately 70% crystalline, have an unusually high tensile strength, are impermeable to water and resistant to chemical and biological attack, and are so stable that they are very difficult to break.

To further enhance cellulose recalcitrance, the glucose repeating unit has six hydrogen bond donors and nine hydrogen acceptors, offering numerous ways of establishing inter- and intramolecular hydrogen bonds. Moreover, due to idiosyncratic arrangements of the pyranose ring and conformational changes of hydroxymethyl groups, cellulose chains form microfibrils which exhibit distinct crystal structures (44–48).

Four distinctive crystalline allomorphs of cellulose have been

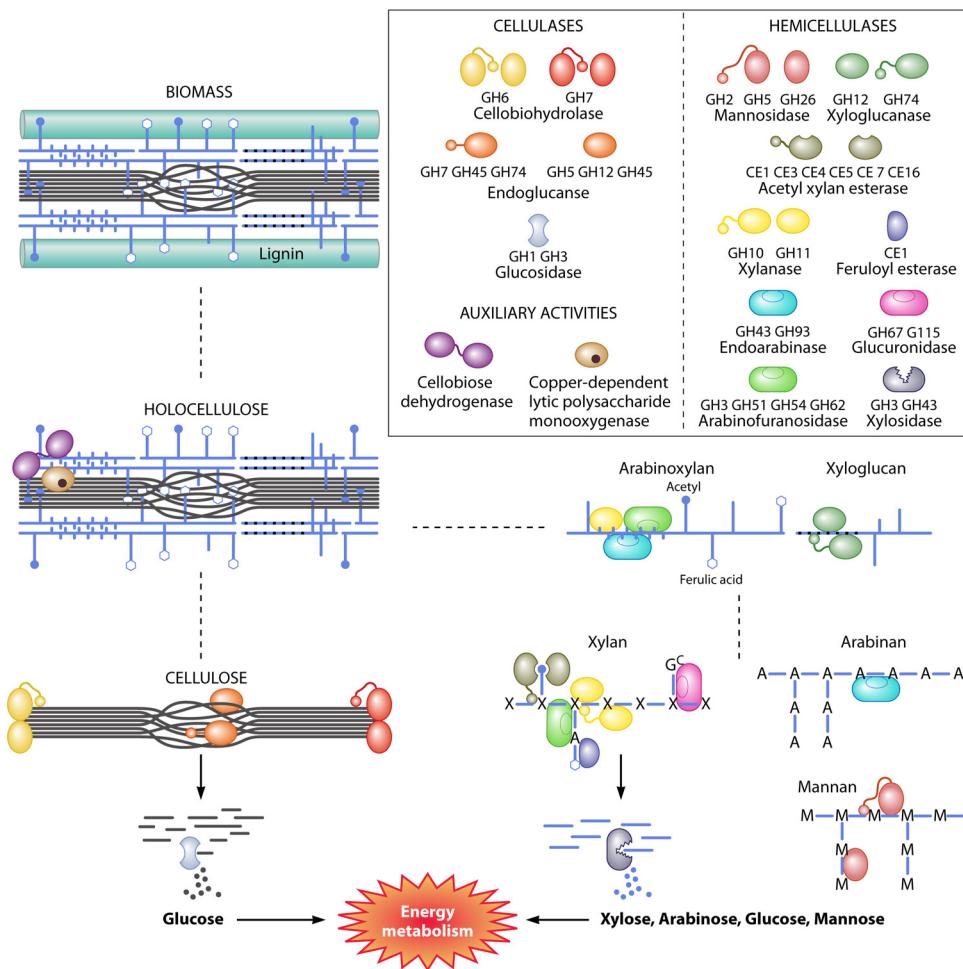


FIG 1 Canonical holocellulose structure and deconstructive hydrolytic enzyme interactions. The main polymers integrating biomass are lignin (boxes) and holocellulose, which includes hemicellulose (light-colored, loosely branched chains) and cellulose (black linear bundled chains). Sugars: X, xylose; A, arabinose; Gc, glucuronic acid; M, mannose. Open hexagons, ferulic acid; closed circles, acetyl groups. Biomass is the principal carbon sink on earth and recruits numerous enzymes to deconstruct cellulose and hemicellulose to sway the carbon cycle via the central energy metabolism. Enzymes needed to deconstruct holocellulose include the following: cellulases, i.e., cellobiohydrolases and endoglucanases, with and without CBMs, β -glucosidases, copper-dependent lytic polysaccharide monooxygenases (LPMOs), and cellobiose dehydrogenases; and hemicellulases, i.e., xylanases, mannosidases, xyloglucanases, xylan acetyl esterases, feruloyl esterases, arabinanases, glucuronidases, and arabinofuranosidase xylosidases.

recognized based on X-ray diffraction patterns and solid-state ^{13}C nuclear magnetic resonance (NMR) spectra (49–51).

Cellulose I, the most abundant form found in nature, is important for bioenergy production and is a mixture of two distinct crystalline forms: cellulose I α , a triclinic one-chain unit with parallel chain stacks with progressive shear parallel to the chain axis, the predominant form isolated from bacteria and fresh water algae (52); and cellulose I β , a monoclinic two-chain unit with parallel cellulose chains stacked with alternating shear, the major form in

higher plants, such as cotton, wood, sugar cane, sorghum, switch grass, ramie, and animal cellulose (53). Cellulose II is the most crystalline, thermodynamically stable form and is a monoclinic unit with antiparallel chains (parallel in cellulose I β) forming a highly rigid macromolecule due to the presence of a three-dimensional (3-D) hydrogen bond network in addition to the C—O—C bonds between the glucopyranose rings. In the absence of such hydrogen bond networks, the chains become flexible (45). Cellulose III I_{D} , III II_{D} , and IV are structural variations which appear when cellulose II or I is treated with chemicals such as ammonia and/or subjected to high-temperature treatments (54–56).

The second major polysaccharide fraction is hemicellulose ($\sim 30\%$), a carbohydrate component loosely defined as polymers extractable by alkaline solutions. Typical alkali-soluble polymers are xylans, mannans, arabinoxylan, xyloglucans, and pectins that contain, in addition to D-glucose, other hexoses (D-mannose, D-galactose, D-fucose, L-rhamnose, and D-galacturonic acid) and pentoses (D-xylose and L-arabinose) (57, 58). Examples of hemi-

TABLE 1 Polysaccharide composition of energy crops and wood

Energy crop	% Sugars in juice	% Biomass in bagasse			Reference
		Cellulose	Hemicellulose	Lignin	
Sugar cane	9.8	43	24	22	123
Sweet sorghum	11.8	45	27	21	123
Hardwood	38–51	17–38	27–32	171	
Softwood	33–42	22–40	21–31	171	

cellulosic polysaccharides are xylans, with a linear xylose β -1,4-linked backbone with few side chain substitutions, and xyloglucan, a β -1,4-linked glucan chain to which xylose residues are bound via α -1,6-glycosidic linkages.

Four main types of backbones are common in hemicelluloses (58, 59). Xylans possess a β -1,4-D-xylopyranose backbone, with glucuronoxylan and arabinoxylan being variants found in plant cell walls. Xyloglucans contain a β -1,4-D-glucopyranan backbone decorated with α -D-xylopyranose residues at position 6 and form hydrogen bonds with cellulose microfibrils. Mannans are polymers possessing a β -1,4-D-mannopyranose backbone, whereas galactomannans and glucomannans are variants that are part of the secondary cell wall of softwoods and, finally, β -glucans are β -(1,3-1,4)-D-glucans found in cereal grains.

Pectins are also extracted with alkali and contain galacturonic acid, rhamnose in backbone structures, and apiose, galactose, and arabinose as side chains, but they are abundant in fruits and not in biomass, such as wood, sorghum, sugar cane bagasse, or corn stover (60).

The third most abundant polymer present in biomass is lignin (\sim 30%), a polymer that surrounds secondary cell walls, resulting in lignified tissues creating a structure that provides mechanical support (e.g., branches and twigs of trees or stems of herbaceous plants). Lignin is formed by polymerization of phenyl propane derivatives, i.e., coumaryl, coniferyl, and sinapyl alcohols, resulting in a solid polymeric structure primarily linked by ether bonds and chemically (covalent bonds) interlinked with hemicellulose (61, 62).

Hemicellulose molecules are cross-linked by phenols, the most abundant of which are *para*-coumaryl and feruloyl acids. Complexes of feruloylated xyloglucan and a *p*-coumaroylated arabinoxylan have been isolated (63, 64), and feruloylated α -1,5-linked arabinan and β -1,4-linked galactan have also been observed (64). The pulp of spruce and pine wood yielded lignin carbohydrate β -1,4-D-galactan complexes (61). Interestingly, a small amount of arabinose was also found in lignin-carbohydrate complexes not associated with arabinoxylan (65), thus suggesting cross-linking via ferulic and/or *p*-coumaric esters to arabinogalactan, α -1,5-linked arabinan, and β -1,4-linked galactan (65).

In summary, cellulose is a linear polymer that is crystalline, strong, and resistant, while hemicellulose is branched and amorphous, with little strength, and pectin along with lignin amalgamates the plant cell wall matrix.

Nearly all fungi and other microorganisms, such as bacteria and archaea, produce enzymes that at least partially degrade plant cell wall polysaccharides (66, 67). They deconstruct cellulose, hemicellulose, lignin, and pectin into simple sugars and phenols, which upon assimilation feed the central energy metabolism to produce chemical energy (ATP) through oxidation, releasing carbon dioxide and water and thus completing the carbon cycle, whereupon carbon dioxide can again be fixed through photosynthesis to produce sugars and plant cell walls.

Figure 1 is a schematic that depicts all significant components of plant cell walls, along with known fungal enzymes adequate for whole conversion of holocellulose (cellulose and hemicellulose) into simple sugars, which are easily metabolized further by fungi and other microorganisms into bioproducts.

The carbohydrate-active enzymes (CAZy) database collectively compiles and assigns into families glycoside hydrolase (GH)-, glycosyl transferase (GT)-, polysaccharide lyase (PL)-, carbohydrate esterase (CE)-, and auxiliary activity (AA)-encoding genes, ac-

cording to a classification system developed by Henrissat and co-workers (68, 69), based on amino acid sequence similarity, secondary and tertiary fold conservation, and stereochemical architecture of catalytic mechanisms, i.e., inversion or retention of the anomeric configuration (68, 70). According to Jovanovic and coworkers (71), only 22 of 114 GH families are critical for biomass decomposition, and 20 are populated with genes from filamentous fungi, including those encoding endo- and exo-acting cellulases, hemicellulases, backbone-degrading and debranching enzymes, and glucosidases (71).

In the historical view of cellulose degradation, only three types of cellulases are required (Fig. 1): cellobiohydrolases that attack the nonreducing (CAZy family GH6) or reducing (GH7) end of a cellulose chain, endoglucanases (GH5 to GH9, GH12, GH44, GH45, GH48, GH51, GH61, and GH124) that cleave internal linkages of cellulose molecules, producing cellobiose, and β -glucosidases (GH1 and GH3) that cleave cellobiose into two glucose molecules. However, some of the “weak” endoglucanases formerly classified as GH61 enzymes were recently recognized to function as copper-dependent lytic polysaccharide monooxygenases (LPMOs) along with cellobiose dehydrogenase, which breaks down cellulose via an oxidative reaction, and therefore were regrouped into a new family: the auxiliary activity family 9 (AA9) for LPMOs (15, 72–75).

For hemicellulose backbone degradation, at least nine different enzymes are needed, depending on the plant's specific cell wall composition (Fig. 1): xylanases (GH10 and GH11) cleave xylose glycoside linkages (76) from xylan homopolymers (77); arabinofuranosidases (GH3, GH51, GH54, and GH62) promote hydrolysis of α -1,2-, α -1,3-, and α -1,5-L-arabinofuranosidic bonds in arabinoxylan, arabinan, and other arabinose-containing hemicelluloses (78); alpha-glucuronidases (GH67 and GH115) hydrolyze the ester linkage between the 4-O-methyl-D-glucuronic acid of glucuronoxylan and lignin alcohols (79); xyloglucanases hydrolyze glycosidic bonds of branched (GH74) or unbranched (GH12) glucose residues of xyloglucan (80); arabinanases (GH43 and GH93) hydrolyze α -L-arabinofuranoside linkages of α -1,5-L-arabinan and release α -1,5-L-arabinobiose from the nonreducing end (81); mannosidases or mannanases (GH2, GH5, and GH26) catalyze random hydrolysis of β -1,4-mannosidic linkages of mannan (82); acetyl xylan esterases (CE1, CE3 to CE5, CE7, and CE16) hydrolyze the O-acetyl substituent present in xylan backbones (83); feruloyl esterases (CE1) catalyze the cleavage of ester bonds between plant cell wall polysaccharides and phenolic acid, mainly ferulic acid (84); and xylosidases (GH3 and GH43) hydrolyze successive xylose residues from the nonreducing end of xylo-oligomers (85).

ENZYME STRUCTURE-FUNCTION AND SUBSTRATE RELATIONSHIPS

Microbial gene models encoding plant cell wall-degrading enzymes exist abundantly in bacteria and fungi. Genes appear as multiple copies (duplications or multiple sequential acquisitions) in the genomes of sequenced microorganisms, and enzymes (proteins) appear to be functionally redundant (with different three-dimensional structural folds harboring similar biochemical functions) or have overlapping biochemical functions (illustrated by GH10 and GH11 xylanases) (60, 86). Furthermore, cellulose-degrading enzymes (Table 2), such as cellobiohydrolases and endoglucanases, comprise similar catalytic functions: endoglucanases

TABLE 2 Complete cellulose degradation gene complement^a

Enzyme	CAZy family	Domain	No. of complete gene models								Avg no. of gene models		
			Aspergillus						Reference species			Aspergillus spp.	Reference species
			<i>clavatus</i>	<i>nidulans</i>	<i>niger</i>	<i>terreus</i>	<i>fumigatus</i>	<i>flavus</i>	<i>oryzae</i>	<i>Trichoderma</i> sp.	<i>N. crassa</i>		
Cellbiohydrolase (see Table S1)	GH6	CBM1	4	4	4	4	3	3	3	2	5	4	4
		CBM1	1	1	1	1	1	0	0	1	1	1	1
		No CBM	1	1	1	1	0	1	1	0	2	1	1
		GH7	CBM1	1	1	1	1	0	0	1	1	1	1
	GH7	CBM1	1	1	1	1	1	2	2	0	1	1	1
		No CBM	1	1	1	1	1	2	2	0	1	1	1
		CBM1	1	0	0	1	1	1	1	0	0	0	0
		No CBM	1	1	0	1	1	1	1	1	0	1	1
Endo-β-1,4-glucanase (see Table S2)	GH5	CBM1	6	4	4	8	7	4	4	5	3	5	4
		CBM1	2	0	0	2	2	1	1	0	1	1	1
		No CBM	1	2	3	1	1	1	1	2	0	1	1
		GH7	CBM1	1	0	0	1	1	0	0	1	0	0
	GH12	CBM1	1	1	0	1	1	1	1	0	2	1	1
		CBM1	0	0	0	0	0	0	0	0	0	0	0
		No CBM	1	0	1	3	1	1	1	1	0	1	1
		GH45	CBM1	0	0	0	0	1	0	0	1	0	0
		No CBM	0	1	0	0	0	0	0	0	0	0	0
β-Glucosidase (see Table S5)	GH1	Intra	10	20	16	16	18	22	22	11	10	18	11
		Extra	3	2	1	2	5	2	3	2	1	3	2
	GH3	Intra	0	1	2	1	0	1	0	0	0	1	0
		Extra	5	7	4	4	7	10	8	6	5	6	6
LPMO (see Table S3)	AA9c1	Intra	6	9	7	12	7	7	8	4	12	8	8
		CBM1	1	1	1	1	1	1	1	0	0	1	0
		No CBM	0	0	0	0	0	0	0	0	0	0	0
		AA9c2	CBM1	1	0	1	2	1	0	0	1	1	1
		No CBM	0	2	1	1	0	1	1	0	2	1	1
		AA9c3	CBM1	0	0	0	0	0	0	0	0	0	0
		No CBM	2	3	2	2	2	3	3	1	0	2	1
		AA9c4	CBM1	0	0	0	0	0	0	0	2	0	1
		No CBM	1	1	0	1	1	0	1	0	2	1	1
		AA9c5	CBM1	0	0	1	1	1	1	0	0	1	1
		No CBM	1	2	1	4	1	1	2	2	4	2	3
Cellobiose dehydrogenase (see Table S4)	AA3	Intra	2	2	1	2	2	2	3	0	2	2	1
		CBM	0	0	0	0	0	0	0	0	0	0	0
		No CBM	2	2	1	2	2	2	3	0	2	2	1
Total no. of cellulase gene models			28	39	31	42	37	38	40	22	32	36	27
% Gene models with CBM1			39%	16%	27%	35%	47%	19%	11%	35%	32%		
% Gene models with no CBM1			61%	84%	73%	65%	53%	81%	89%	65%	68%		

^a For complete gene and protein information, refer to Tables S1 to S5 in the supplemental material. Gray-shaded entries are totals.

hydrolyze glucose-glucose glycoside bonds randomly, while cellobiohydrolases cleave the same glycoside bond from terminal ends of cellulose molecules, producing cellobiose. Furthermore, two types of cellobiohydrolases are routinely found: CAZy families GH6 and GH7, which cleave cellobiose from the nonreducing and reducing ends, respectively. Endoglucanases also appear as functional variants, based on the structure relationships of the CAZy classification system (GH5, GH7, GH12, and GH74).

Hemicelluloses (Fig. 1) are structurally complex polymers containing a variety of monomers (xylose, arabinose, glucose, and

glucuronic acid), backbones, and side chains with substituted sugars and attached phenols and therefore require a larger number of distinctive enzymes for complete degradation (see Table 8). Thus, the more recalcitrant a polymer is, even if it is bonded homogeneously, the more redundant functional enzymes are employed, and the more diversified a structural design is, the more functional types with less redundancy are demanded. This apparent gene redundancy and functional diversity related to the substrate are reflected in the genome content of the group of fungi studied here (see Tables 2, 8, and 10).

In order to determine the functional importance of this observed redundancy (of genes) and functional multiplicity (of proteins), we examined the full gene complements of seven *Aspergillus* genomes to precisely define which genes encode functionally similar proteins and to spot genes that appear to be redundant. For example, we show that aspergilli encode exactly two cellobiohydrolases belonging to family GH7, which are structurally and functionally identical at the catalytic domain; however, one contains a carbohydrate-binding module (CBM), while the other does not. On the other hand, we also show that extracellular β -glucosidases are encoded by multiple loci with acquired functional specialization, such as pH and temperature sensitivity as well as susceptibility to inhibitors and activators. As we proceed with the descriptive analysis of the complete holocellulose gene complement, the initial notions of genetic multiplicity (multiple gene copies) and functional redundancy (proteins with similar functions) rapidly disappear once enzymes are organized into their functional and interactive components, and only a few cases of gene duplication remain.

The presented evidence provides strong indications that, over time, fungal genomes acquire multiple gene copies, perhaps through horizontal gene transfer, as shown for other systems (87) that are modified through recombination in order to adapt to the notoriously recalcitrant substrate accessibility of plant cell walls.

Cellulases

Cellulose is a structurally uncomplicated polysaccharide that is recalcitrant toward degradation due to its crystallinity and interconnection with other cell wall polymers, such as hemicellulose and lignin. Thus, even though hydrolysis of cellulose is a simple glucose-glucose glycoside bond rupture, several enzymes are needed to perform this function.

The available three-dimensional structures of cellobiohydrolases (Fig. 2) suggest a tunnel-like conformation around the active site, fitting a single cellulosic chain at the reducing (GH7) or nonreducing (GH6) terminus (88–90). In contrast, endoglucanases (GH5 to -9, GH12, GH44, GH45, GH48, GH51, GH61, and GH124) are shaped by an open groove or cleft, into which a linear amorphous cellulose chain can fit randomly (Fig. 3).

The model *T. reesei* cellulase complement has been studied extensively and comprises four endoglucanases (EGI/Cel7B, EGII/Cel5A, EGIII/Cel12A, and EGV/Cel45A), a lytic polysaccharide monooxygenase (AA9; originally incorrectly identified as a hydrolase and therefore called EGIV/Cel61A), and two cellobiohydrolases (CBHI/Cel7A and CBHII/Cel6A) that act synergistically to break down cellulose into cellobiose (91–95). In addition, two β -glucosidases (BGLI/Cel3A and BGLII/Cel1A) hydrolyze cellobiose into glucose (91, 96–98).

Table 2 describes aspergillus hydrolytic cellulases, cellobiohydrolases, endoglucanases, and β -glucosidases as well as oxidative enzymes, lytic polysaccharide monooxygenases, and cellobiose dehydrogenase.

Cellobiohydrolases. Two cellobiohydrolases, GH6 and GH7, hydrolyze cellobiose from the cellulosic nonreducing and reducing ends, respectively. Enzymes that sequentially remove cellobiose molecules from a cellulose chain are termed processive enzymes (99). Two forms of cellobiohydrolases are present in aspergilli: cellobiohydrolases with and without a CBM (Table 3). Enzymes with CBMs bind tightly to cellulose molecules and aid in the removal of cellobiose molecules from terminal ends, and cel-

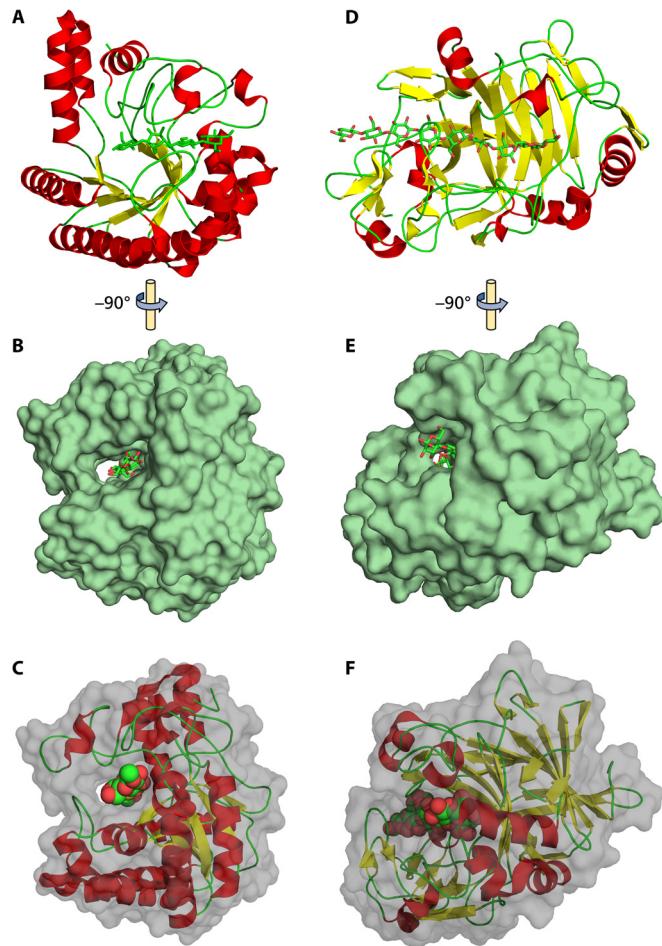


FIG 2 GH6 (A to C) and GH7 (D to F) cellobiohydrolases. (B and E) Typical tunnel-shaped catalytic cleft found in cellobiohydrolases. (C and F) Cleft depth. Cellobiohydrolases fold into an enclosed catalytic core shaped by a β -sandwich with two large, antiparallel β -sheets packed onto each other, forming a long cellulose-binding tunnel (226). The cellulosic substrate chain has to travel through the tunnel, where β -1,4-glycosidic bonds of cellobiose molecules (dimers) are hydrolyzed off the ends (GH6 or GH7 enzymes). The three-dimensional structures are for *Trichoderma reesei* GH6 (CBHII; PDB entry 1QK2) (108) and GH7 (CBHI or Cel7A; PDB entry 4C4C) enzymes (227).

lobiohydrolases lacking CBMs are also present in all surveyed fungal genomes, showing clear processive activity (100, 101).

Cellobiohydrolases (Fig. 2) belonging to family GH6 (EC 3.2.1.91), also known as cellulase family B (Cel6A and CBHII), are thought to hydrolyze β -1,4-D-glycosidic linkages in cellulose and celotetraose, releasing cellobiose from the nonreducing ends of cellulose molecules (102–104). GH6 cellobiohydrolases function through an inversion of anomeric stereochemistry, as indicated by NMR (105) with cellobiohydrolase II from *T. reesei* (*Hypocrea jecorina*). The first 3-D structure of cellobiohydrolase II (*Trichoderma reesei* Cel6A) provided evidence recognizing the catalytic general acid in the inverting mechanism (106–108).

Cellobiohydrolases belonging to family GH7 (EC 3.2.1.176), also known as cellulase family C, are thought to hydrolyze β -1,4-D-glycosidic linkages in cellulose, releasing cellobiose from the reducing ends of cellulose molecules. GH7 enzymes retain two catalytic amino acid residues near the consensus, Glu-X-Asp-X-X-Glu, where the first Glu acts as the catalytic nucleophile and the

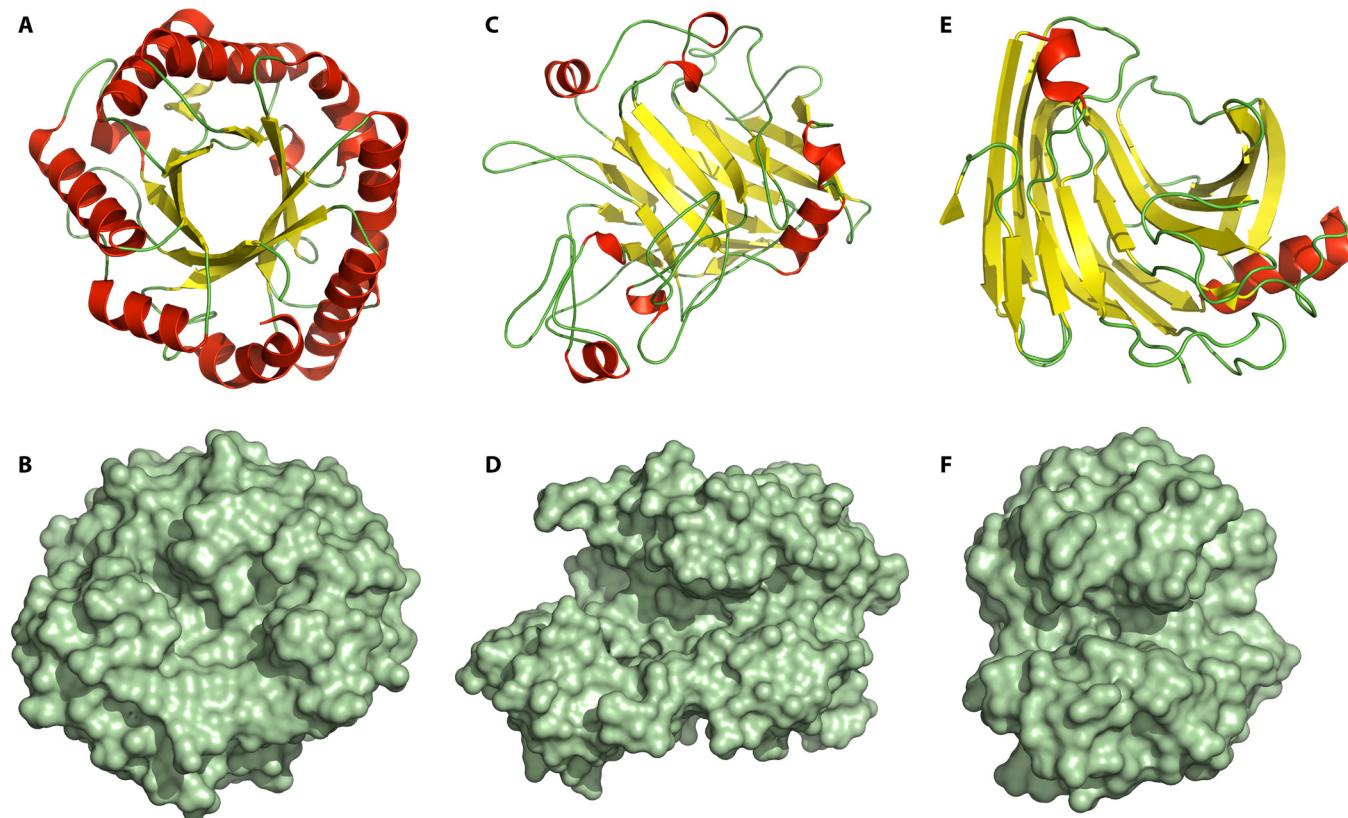


FIG 3 Three-dimensional structures of GH5 (A and B), GH7 (C and D), and GH12 (E and F) endoglucanases. (B, D, and F) Well-defined open clefts in these endoglucanase families. Endoglucanases from the GH5 family show a catalytic module with a typical compact 8-fold β/α barrel architecture, forming an open cleft similar to those of GH7 and GH12 endoglucanases, which share the β -jelly-roll topology with an extended, open substrate-binding groove. Endoglucanases with the open cleft configuration bind randomly to internal portions of a cellulose chain and cleave β -1,4-glycosidic bonds, resulting in shortened fragments. The three-dimensional structures are for the *Thermoascus aurantiacus* GH5 endoglucanase Cel5A (PDB entry 1GZ1), the *Trichoderma reesei* GH7 enzyme EGI (PDB entry 1EG1), and the *Trichoderma reesei* GH12 enzyme EGIII (Egl3 or Cel12A; PDB entry 1H8V) (116, 119, 122).

second Glu as a general acid/base (109, 110). Processivity is the key property of cellobiohydrolases, which employ reducing-end exo- and endo-mode initiation in parallel (88, 90, 111, 112). Intrinsic processivity (absence of obstacles or blocking) is greater than ob-

served processivity, indicating that cellobiohydrolase activity is limited by something blocking the enzyme rather than by the enzyme's intrinsic tendency to fall off a cellulose chain (112).

Two forms of cellobiohydrolase family GH7 (Cbh1 and CelD) are

TABLE 3 Genome-wide distribution of fungal cellobiohydrolases

Fungus	Cellobiohydrolase(s)			
	GH6 (nonreducing end)		GH7 (reducing end)	
	Without CBM	With CBM	Without CBM	With CBM
Strains with a complete cellobiohydrolase set				
<i>A. clavatus</i> NRRL-1	ACLA_025560	ACLA_062560	ACLA_088870	ACLA_085260
<i>A. nidulans</i> FGSC A4	AN1273	AN5282	AN5176	AN0494
<i>A. niger</i> CBS 513.88	ANI_1_1704074	ANI_1_300104	ANI_1_2134064	ANI_1_1574014
<i>A. terreus</i> NIH 2624	ATEG_00193	ATEG_07493	ATEG_03727	ATEG_05002
Strains with an incomplete cellobiohydrolase set				
<i>A. fumigatus</i> Af293			AFUA_6G07070	AFUA_6G11610
<i>A. flavus</i> NRRL-3357	AFLA_069820	AFLA_067550, AFLA_021870		
<i>A. oryzae</i> RIB 40	AOR_1_734074	AOR_1_608164, AOR_1_1654194		
Strains used for reference				
<i>H. jecorina</i> taxid 51453		GUX2_HYPJE		GUX1_TRIRE
<i>N. crassa</i> OR 74A	NCU03996, NCU07190	NCU09680	NCU05104	NCU07340

common to all aspergilli and have structurally similar catalytic domains (Fig. 2). However, only Cbh1 contains a CBM that binds to cellulose. Both enzymes are catalytically functional, but CelD does not have a CBM and has a four-amino-acid deletion on the tunnel-obstructing loop, providing a continuous opening in the absence of a CBM (100). The fact that only Cbh1 binds to the substrate and in combination with CelD exhibits strong synergy only when Cbh1 is present in excess suggests that Cbh1 unties enough chains from cellulose fibers to enable processive access of CelD (100). Interestingly, a direct correlation of water content of the substrate and the absence or presence of CBMs on cellobiohydrolases was established, suggesting that, in nature, mixing and matching of catalytic domains with cellulose-binding domains result in functional interactions that optimize the catalytic output (101).

Endoglucanases. Endoglucanases hydrolyze β -1,4-glycoside bonds intramolecularly and randomly along the noncrystalline portion of cellulose molecules. Four CAZy families are present in aspergilli: GH5, GH7, GH12, and GH45. Most CAZy family endoglucanases show one gene per family, sometimes one with and another without a CBM. Some multiplicity is also observed in GH12 and GH5 (Table 2).

GH5 endoglucanases (Fig. 3), formerly known as cellulase A, are the main cellulases in fungi and other organisms as well (113, 114). They hydrolyze β -1,4-D-glycosidic bonds randomly and internally of the amorphous region of cellulose. The catalytic core domain of Cel5A from *T. reesei* determined at 2.05 Å shows a substrate-binding pocket consisting of a deep catalytic cleft within a shallow groove, consistent with other structural studies of GH5 endoglucanases (115). The *Thermoascus aurantiacus* GH5 endoglucanase, which consists of a catalytic module with compact 8-fold β/α -barrel architecture (116), has a long, tryptophan-rich substrate-binding groove suggesting substrate-binding subsites at positions -4 to $+3$, in addition to the two conserved catalytic glutamates (116).

GH7 endoglucanases (Fig. 3B) are similar to GH16 enzymes found in plants and to bacterial agarases, cleaving β -1,3- and/or β -1,4-glycosidic linkages. Members are related by amino acid sequence similarity, the retaining hydrolytic mechanism, and catalytic residue identity (117). These endoglucanases share the β -jelly-roll topology and the retaining catalytic mechanism (118–120). The endoglucanase catalytic core domain from *T. reesei*, determined at 3.6-Å resolution, reveals an extended, open substrate-binding cleft rather than a tunnel like the one found in cellobiohydrolases (119), showing that the tunnel-forming loops have been deleted, which results in an open active site enabling random and internal binding (119).

GH12 endoglucanases (Fig. 3C) form a β -jelly-roll structure with two β -sheets, of six and nine strands, packed on top of one another, and one α -helix. The concave surface forms a substrate-binding groove in which the active site residues, a carboxylic acid trio similar to those of GH7 and GH16 glycoside hydrolases, are located (121, 122). The GH12 endoglucanase from *A. niger* has also been crystallized and follows a similar three-dimensional architecture, even though the amino acid sequence similarity is somewhat less conserved (123).

Phylogenetic, functional, and substrate specificity analyses of 30 endoglucanases belonging to six GH families (GH5, GH6, GH7, GH9, GH12, and GH45) suggest a structure-function relationship based on active site conformation and the catalytic mechanism (124). Moreover, GH5 endoglucanases are part of a large

family that can be grouped into subfamilies (over 31), while all aspergillus endoglucanases fall into two GH5 subfamilies (subfamilies 5 and 7) showing little structural variation (125).

Table 4 shows that compared to cellobiohydrolases, endoglucanases associate least often with CBMs, although ~30% of all observed endoglucanases are linked to a CBM. On average, the aspergillus genome encodes one or two GH5 endoglucanases with a CBM and one or two without and at least one GH12 endoglucanase with no CBM. GH7 endoglucanases appear to be intermediate, as they are sometimes partially represented or absent, and GH45 endoglucanases are rare. There seems to be some redundancy (same gene model) of endoglucanases with similar structural folds and catalytic activities, especially among families GH5 and GH12. However, whether these apparently redundant gene copies have diversified over time and the encoded proteins acquired novel specialized functions, such as differentiated pH and temperature sensitivity or inhibition and/or activation sites, remains unknown.

Copper-dependent lytic polysaccharide monooxygenases. The recently created auxiliary activity family 9 (AA9), formerly included in glycoside hydrolase family 61, is in fact comprised of copper-dependent lytic polysaccharide monooxygenases (LPMOs).

The entire family was originally classified as glycoside hydrolase family 61 because the original protein expressed in yeast had some endoglucanase activity and was regulated similarly to other *T. reesei* cellulase genes (94). However, GH61 (AA9) enzymes significantly enhance cellulolytic activity on lignocellulosic substrates in combination with other cellulases, due to their action on the cleavage of cellulose chains through an oxidative process of C-1, C-4, and C-6 carbons (126). LPMOs use copper as the catalytic metal, oxygen, and reducing agents for their activity (126). The secretome of *Thielavia terrestris* contains six AA9 proteins, comprising 10% of the total soluble protein secreted into the medium of cultures grown in the presence of cellulose (72). Thus, the classical hydrolytic mechanism for the degradation of plant polysaccharides was recently challenged by the landmark discovery (15, 75, 127) that oxidoreductase systems, such as fungal LPMOs, directly oxidize cellulose, breaking glycoside bonds and generating aldones and lactones (15, 75, 127–129).

The first AA9 LPMO high-resolution structural model was derived from *T. reesei* (129, 130). The protein core is a twisted β -sandwich built up of nine β -strands forming a compact single-domain β -sandwich with a large buried ionic network (Fig. 4). A functional metal-ion-binding site is coordinated by three conserved histidines located at the surface near the N terminus, and the requirement for a divalent metal ion for catalytic activity was determined experimentally (74, 75). There is a noteworthy structural similarity between AA9 LPMOs and the chitin-binding protein CBP21 (from *Serratia marcescens*), a protein that stimulates the chitin-degrading activity of chitinases, with no hydrolytic activity itself (130). Notably, alignment of GH61 endoglucanases with other glucanases did not support the positioning of the two conserved catalytic acidic residues that are present in the catalytic site of all glycoside hydrolases, supplying the acid component that is indispensable for hydrolysis of glycoside bonds (72).

Because of the large number of potential LPMOs (formerly GH61 endoglucanases), we performed a neighbor-joining phylogenetic tree analysis in order to group similar proteins into five distinctive clans. Table 5 shows all AA9 LPMOs of aspergilli grouped into five clans (c1

TABLE 4 Genome-wide endoglucanase content in aspergilli^a

Fungus	Presence of CBM1	Endo- β -1,4-glucanase(s)				No. of cellulases per genome
		GH5	GH7	GH12	GH45	
		± CBM	Total			
<i>A. clavatus</i> NRRL1	No	ACLA_081650		ACLA_066030	ACLA_007820	3
	Yes		ACLA_085250, ACLA_081310		ACLA_098940	3
<i>A. flavus</i> NRRL3357	No	AFLA_087870		AFLA_118170	AFLA_138380	3
	Yes		AFLA_045290			1
<i>A. fumigatus</i> Af293	No	AFUA_6G07480		AFUA_7G01540	AFUA_7G06150	3
	Yes		AFUA_6G11600, AFUA_2G09520	AFUA_6G01800		4
<i>A. nidulans</i> A4	No	AN1285, AN5214		AN3418		AN6786
	Yes					4
<i>A. niger</i> CBS513.88	No	ANI_1_916144, ANI_1_3102014, ANI_1_1120064			ANI_1_398124	4
	Yes					0
<i>A. oryzae</i> RIB40	No	AOR_1_2698174		AOR_1_514024	AOR_1_194014	3
	Yes		AOR_1_1216054			1
<i>A. terreus</i> NIH2624	No	ATEG_03677		ATEG_08700	ATEG_09894, ATEG_05519, ATEG_07420	5
	Yes		ATEG_05003, ATEG_04390	ATEG_08705		3
<i>Hypocrea jecorina</i>	No	AAP57754/cel5b, ABA64553			BAA20140/egl	3
	Yes			P07981/egl1		2
<i>N. crassa</i> OR74A	No			NCU05057, NCU04854		2
	Yes	NCU00762				1

^a Among the aspergilli, there were 18, 9, 9, and 6 cellulases in the GH5, GH7, GH12, and GH45 families, respectively, for a total of 37 cellulases. ±, with or without.

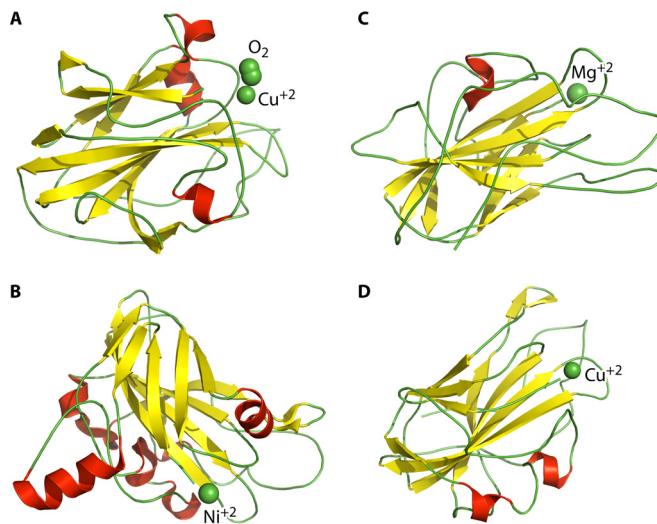


FIG 4 Lytic polysaccharide monooxygenases (LPMOs). LPMOs, which are classified in the AA9 family (formerly GH61), are bivalent ion-dependent lytic polysaccharide monooxygenases. These proteins cleave cellulose chains with oxidation of various carbons (C-1, C-4, and C-6). The LPMO three-dimensional structures are for *Neurospora crassa* (PDB entry 4EIR) (129) (A), *Hypocrea jecorina* (PDB entry 2VTC) (130) (B), *Thielavia terrestris* (PDB entry 3EJA) (72) (C), and *Phanerochaete chrysosporium* (PDB entry 4B5Q) (126) (D).

to c5), based on phylogenetic neighbor-joining relationships (see Table S3 in the supplemental material).

Aspergilli, in general, have an average of 8 AA9 LPMOs, whereas *A. clavatus* has 6, *A. niger* 7 and *A. terreus* 12. Aspergilli possess one AA9c1, one or two AA9c2, two or three AA9c3, one AA9c4, and several AA9c5 LPMOs. Most AA9 LPMOs are not linked to CBMs, except for AA9 clan 1, in which every enzyme is linked to a CBM. Thus, AA9 LPMOs are a diverse group of enzymes whose activity remains to be elucidated fully. The interesting aspect of this group of enzymes is the fact that they apparently do not work alone; they interact with other enzymes, such as cellobiose dehydrogenase (CDH) (15, 127), enhancing overall cellulolytic activity.

Cellobiose dehydrogenase. Cellobiose dehydrogenase (EC 1.1.99.18) is an enzyme that oxidizes cellobiose to cellobiolactone in the presence of an electron acceptor, such as cytochrome *c*, dichlorophenol-indophenol, or ferricyanide, producing cellobio-1,5-lactone and a reduced acceptor (131, 132). Cellobiose dehydrogenases are well studied in white and brown rot and plant-pathogenic as well as composting fungi from the dikaryotic phyla *Basidiomycota* and *Ascomycota* under cellulolytic culture conditions (133).

CDH is the major oxidoreductase secreted by some fungi (but not aspergilli) growing on biomass that contains cellulose, and it catalyzes the oxidation of cellobiose and longer cellobextrins to

TABLE 5 Fungalolytic polysaccharide monooxygenases of the AA9 family

Fungus	Presence of LPMO(s)					No. of gene models in a clan	LPMOs not categorized
	CBM1	Clan 1	Clan 2	Clan 3	Clan 4		
<i>A. clavatus</i> NRRL1	No						
	Yes	ACLA_059790	ACLA_060890				6
<i>A. fumigatus</i> AF293	No						
	Yes	AFLA_077840					7
<i>A. nidulans</i> A4	No						
	Yes	AFUA_8G06830	AFUA_3G03870				
<i>A. niger</i> CBS513.88	No						
	Yes	ANI602	AN3511, AN9524				9
<i>A. oryzae</i> RIB40	No						
	Yes	ANL_1_576104	ANL_1_748074, ANL_1_2148184				7
<i>A. terreus</i> NIH2624	No						
	Yes	AOR_1_3066174	AOR_1_148194 AOR_1_380164, AOR_1_4064, AOR_1_1358144				8
	Yes	ATEG_07790	ATEG_07920, ATEG_04210				
Total no. of gene models	7	11	17	5	16	56	
Other							
<i>H. jecorina</i>	Yes	CAA71999					
	No						
<i>T. auranticius</i>	No						
<i>T. terrestris</i>	No						
<i>T. terrestris</i>	No						
<i>N. crassa</i> OR74A	No						
	Yes	NCU07520, NCU05969					
	Yes	NCU07760					
Total no. of gene models	0	4	1	4	7	19	

TABLE 6 Fungal cellobiose dehydrogenases

Fungus	Cellobiose dehydrogenase(s)	No. of CDH genes per genome
<i>A. clavatus</i> NRRL1	ACLA_076510, ACLA_094490	2
<i>A. flavus</i> NRRL3357	AFLA_001890, AFLA_023820	2
<i>A. fumigatus</i> Af293	AFUA_2G17620, AFUA_2G01180	2
<i>A. nidulans</i> A4	AN7230.2, AN3962	2
<i>A. niger</i> CBS513.88	ANI_1_168174	1
<i>A. oryzae</i> RIB40	AOR_1_98134, AOR_1_712114, AOR_1_2566154	3
<i>A. terreus</i>	ATEG_09993, ATEG_08150	2
<i>H. jecorina</i>		0
<i>N. crassa</i> OR74A	NCU05923, NCU00206	2

1-5- δ -lactones (134). CDH enhances crystalline cellulose degradation by coupling the oxidation of cellobiose to reductive activation of copper-dependent polysaccharide monooxygenases that catalyze the insertion of oxygen into C—H bonds adjacent to the glycoside linkage (15). Deletion of *cdh-1*, the gene that encodes the major *N. crassa* CDH, resulted in reduced cellulase activity, and addition of purified CDH from *M. thermophila* to the $\Delta cdh-1$ strain resulted in a 1.6- to 2.0-fold stimulation of cellulase activity (15). These results suggest that CDH acts as a cofactor for LPMO enzymes. This may (or may not) be true in nature, because redox agents, such as gallate or ascorbate (135), can substitute for CDH.

Lactones hydrolyze spontaneously in solution or are hydrolyzed enzymatically by lactonases to generate aldonic acids (136). Oxidation of cellobiose takes place in the flavin domain following electron transfer to the heme domain, and the reduced heme is able to reduce a wide variety of substrates, including quinones, metal ions, and organic dyes. Reduced cellobiose dehydrogenase can also react with molecular oxygen and interact with the newly discovered copper-dependent lytic polysaccharide monooxygenases that directly oxidize crystalline cellulose (15, 137). The current hypothesis for the function of cellobiose dehydrogenase involves the generation of hydroxyl radicals, formed via reduction of an extracellular ferric complex (138, 139) that takes part in Fenton chemistry, with hydrogen peroxide produced by CDH transferred to an array of acceptor oxidases, such as LPMOs and others that remain unknown (128, 140).

Phylogenetic analysis of all known cellobiose dehydrogenase-encoding genes showed a separation into three classes: class I, found only in *Basidiomycota*; class II, found in *Ascomycota*, frequently fused to a CBM (133, 139); and class III, found only in *Ascomycota*, where they lack a CBM (141).

Cellobiose dehydrogenases are typically monomeric proteins consisting of two domains joined by a protease-sensitive linker region (142). Domains found in cellobiose dehydrogenase, such as heme-binding cytochrome, GMC oxidoreductase, Rossmann-fold NAD(P) $^{+}$ binding, choline dehydrogenase, and flavoprotein (BetA) domains, are highly conserved among all aspergilli (Table 6).

In general, aspergilli (Table 6) contain two CDHs, whereas *A. niger* has one and *A. oryzae* three. CDHs are not linked to a CBM. To date, all investigated CDHs have been reported to bind specifically to cellulose. For some CDHs, the cellulose-binding ability is attributed to a divergent carbohydrate-binding module (143),

whereas other types of CDHs, with no cellulose-binding domain at all, bind to the cellulose surface through the FAD-binding domain (144).

β -Glucosidases. β -Glucosidases are key enzymes in lignocellulosic hydrolysis because they convert cellobiose and other cellobiosaccharides into glucose (145). β -Glucosidases are classified by CAZy into two families: GH1 and GH3 (69, 146, 147). The structure of a GH1 β -glucosidase from *T. reesei* (TrBgl2; PDB entry 3AHY) has a classical $(\alpha/\beta)_8$ -TIM barrel fold. The active site of TrBgl2 consists of a 15- to 20-Å-deep slot-like cleft located on connecting loops at the C-terminal end of the β -sheets of the TIM barrel (148).

The crystal structure for a GH3 β -glucosidase from *A. aculeatus* has been published (PDB entry 4IIB-H) (149). AaBGL1 is a dimer in solution, and the monomer consists of three domains: a catalytic TIM barrel-like domain, an α/β sandwich domain, and an FnIII (fibronectin type III) domain. Linkers connect these domains.

The genomes of seven aspergilli encode extracellular β -glucosidases belonging to two CAZy families: GH1 and GH3. Table 7 shows that genes for extracellular GH1 β -glucosidases are scarce, averaging 1 gene per genome, while genes for GH3 glucosidases are widely abundant, averaging 7 gene models per genome, with *A. flavus* harboring 11 and *A. fumigatus* harboring 5. On average, aspergilli encode 18 β -glucosidases per genome, with 9 being extracellular and 9 intracellular (see Table S5 in the supplemental material). Intracellular β -glucosidases are distributed between the GH1 and GH3 families similarly to extracellular ones, with aspergilli carrying, on average, 2 and 7 gene models, respectively (see Table S5). The β -glucosidase gene model content per *Aspergillus* genome varies from 10 to 22 (see Table S5), suggesting duplications or sequential acquisition events for similar gene models within this group of fungi.

***Aspergillus* cellulose degradation summary.** Aspergilli contain, on average, 36 cellulase gene models (Table 2), while *H. jecorina* and *N. crassa* contain an average of 27, that can be divided into two main functional categories: hydrolytic (27 gene models per average genome) and oxidative (10 gene models per average genome).

Hydrolytic enzymes are abundant and can be grouped further by the way they interact with the substrate, the three-dimensional shape of the catalytic domain, and the type of hydrolysis they promote (retaining or inverting).

Cellobiohydrolases interact with the reducing (average of 3 per genome) or nonreducing (average of 3 per genome) terminus of crystalline cellulose fibers, splitting off cellobiose, occasionally linked to a CBM (average of 1 per genome) or without a CBM (average of 2 per genome).

Endoglucanases interact randomly with amorphous cellulose through an open-cleft catalytic domain that appears in three structural folds: the $(\alpha/\beta)_8$ (GH5; averages 2 gene models per genome), jelly roll (GH7 and GH12; average 1 gene model [each] per genome), and 7-fold propeller (GH74; averages 1 gene model per genome) structural configurations. Endoglucanases bind internally to cellulose molecules and hydrolyze glycoside bonds through a retaining (GH5, GH7, and GH12; average 1 gene model [each] per genome) or inverting (GH74; averages 1 gene model per genome) mechanism, thus exponentially multiplying cellulose termini accessible to cellobiohydrolases and β -glucosidases.

Oxidative enzymes include the major cellobiose dehydrogenase (average of 2 gene models per genome), which produces ox-

TABLE 7 Fungal β-glucosidases

Fungus	Domain type	β-Glucosidase(s) or no. of glucosidase genes		No. of glucosidase genes per genome	
		GH1	GH3		
<i>Aspergillus</i>					
<i>A. clavatus</i> NRRL1	Extra		ACLA_010340, ACLA_087610	2	10
	Intra	3	5	8	
<i>A. flavus</i> NRRL3357	Extra	AFLA_090730	AFLA_063040, AFLA_014190, AFLA_069670, AFLA_023350, AFLA_051140, AFLA_011000, AFLA_128480, AFLA_057310, AFLA_012950	10	22
	Intra	2	10	12	
<i>A. fumigatus</i> Af293	Extra		AFUA_7G06140, AFUA_1G17410, AFUA_6G08700, AFUA_8G02100, AFUA_1G05770, AFUA_5G07080	6	18
	Intra	5	7	12	
<i>A. nidulans</i> A4	Extra	AN9183	AN10482, AN7396, AN2828, AN3904, AN5976, AN7915, AN4102, AN3949, AN2217, AN1804	11	20
	Intra	2	7	9	
<i>A. niger</i> CBS513.88	Extra	ANI_1_1704094, ANI_1_1736184	ANI_1_1526134, ANI_1_2196074, ANI_1_2010064, ANI_1_2160064, ANI_1_456164, ANI_1_2122094, ANI_1_24094, ANI_1_380154, ANI_1_274124	11	16
	Intra	1	4	5	
<i>A. oryzae</i> RIB40	Extra		AOR_1_362074, AOR_1_230194, AOR_1_706074, AOR_1_226034, AOR_1_1470114, AOR_1_2662154, AOR_1_586184, AOR_1_504114, AOR_1_958164, AOR_1_6194, AOR_1_236014	11	22
	Intra	3	8	11	
<i>A. terreus</i> NIH2624	Extra	ATEG_02657	ATEG_07931, ATEG_10320, ATEG_07419, ATEG_10274, ATEG_02713, ATEG_08027, ATEG_06617, ATEG_09314, ATEG_03047	10	18
	Intra	2	6	8	
Total glucosidases in GH family		18	47		126
<i>Hypocreahojecorina</i>	Extra		AAP57755, EGR50829, AAP57760	3	11
	Intra	2	6	8	
<i>N. crassa</i> OR74A	Extra		NCU04952, NCU03641, NCU08755, NCU09923	4	10
	Intra	1	5	6	

xygen radicals utilized by lytic polysaccharide monooxygenases (average of 8 gene models per genome) to oxidize glycoside linkages, breaking long crystalline cellulose chains to make them accessible to other hydrolytic enzymes.

Hemicellulases

While cellulose is a simple linear polymer that is recalcitrant to enzymatic digestion, hemicellulose is a heterogeneous mix of

polymers, with linear and branched sections composed of various types of sugars and decorated by side groups. Thus, hemicellulases comprise a diverse group of enzymes which catalyze hydrolysis of sugar bonds and side chains, such as ferulic acid and acetyl groups (Fig. 1). While cellulose is invariable in structure among plants, hemicellulose composition varies significantly between plant species. Thus, the makeup of hemicellu-

lose in biomass, e.g., corn stover versus sugar cane bagasse, varies considerably.

Hemicellulases can be grouped into the following functional types: enzymes that hydrolyze backbones, enzymes that remove side chains, and accessory enzymes that remove decorations such as acetyl groups.

Hemicellulose backbone-hydrolyzing enzymes. Backbone-hydrolyzing hemicellulases comprise all the enzymes needed to degrade the main polymer of xylan, mannan, and arabinan, as well as mixed structures, such as arabinoxylan, glucoronoxylan, xyloglucan, galactomannans, and glucomannans, composed largely of xylose, mannose, and arabinose monomers.

(i) **Endo-1,4- β -xylanases.** Xylanases are defined as enzymes that promote hydrolysis of β -1,4-D-xylosidic linkages. Xylanases belonging to the GH10 and GH11 families are retaining enzymes with stereoselective hydrolysis of xylan or β -xylobiosides through a double-displacement mechanism involving a covalent xylobiosyl-enzyme intermediate (150–152). The double-displacement mechanism involves a glycosyl-enzyme intermediate, which is formed and hydrolyzed with general acid/base catalytic assistance, usually involving a Glu (153). Some xylanases promote incomplete hydrolysis of xylan substrates, resulting in xylobiose, xylotriose, and xylotetraose (154, 155).

GH10 xylanases exhibit 8-fold TIM barrel [$(\beta/\alpha)_8$] structures containing a deep active site groove, consistent with the endo mode of action (156). GH11 xylanases exhibit β -jelly-roll folding structures with two β -sheets and an α -helix, resembling a partially closed right hand (157).

Table 8 shows that *A. niger* has a single GH10 but four GH11 xylanases. All other aspergilli have more than one GH10 or GH11 xylanase gene, averaging three genes per genome. Some of the GH10 xylanases are linked to a CBM, i.e., *A. clavatus*, *A. flavus*, *A. fumigatus*, and *A. terreus* have one gene each, and none of the GH11 xylanases are linked to a CBM. Thus, it seems that xylanases have been acquired multiple times by most fungal genomes, and both structural types (GH10 and GH11) are represented in all analyzed fungi. **Table 8** shows that the presence of a CBM is only occasionally found in GH10 xylanases, while the vast majority of GH10 and GH11 xylanases are not linked to a CBM.

Most known xylanases are grouped into the GH10 and GH11 families, although a few bacterial xylanases were recently characterized in detail and ascribed to the GH5 family (158). We did not find GH5 xylanases in aspergilli, although the CAZy database moved some GH5 members into GH30, and we found one characterized GH5 xylanase from *T. reesei* (159), which had only two gene models that may be homologous in aspergilli. However, the amino acid sequence homology was not clear. Since none of the other aspergilli showed GH5 xylanases, we did not include them in our tables.

(ii) **β -Xylosidases (4- β -D-xylan xylohydrolases).** β -Xylosidases hydrolyze β -1,4-D-xylans, oligomers, and xylobiose to remove successive single D-xylose residues from the nonreducing terminus. GH3 xylosidases follow xylanases, with a retaining mechanism of hydrolysis, while GH43 xylosidases are inverting enzymes (160). *Talaromyces emersonii* (β -XTE) and *T. reesei* (β -XTR) hydrolyze xylobiose, producing β -D-xylose (160). In *Aspergillus awamori*, the enzymatic hydrolysis of *p*-nitrophenyl β -D-xylopyranoside occurs with overall retention of the substrate anomeric configuration, suggesting cleavage of xylosidic bonds through a double-displacement mechanism (161, 162). The catalytic amino acid residues of the extracellular β -D-xylosidase,

which hydrolyzes *p*-nitrophenyl β -D-xyloside as a substrate, involve a carboxylate and a protonated group part for binding of the substrate; however, only a carboxylate group is needed for cleavage (163).

Only GH43 xylosidases have been crystallized, and they display a five-blade β -propeller fold (164, 165). This three-dimensional structural fold also applies to other GH43 members, such as arabinoxylan arabinofuranohydrolases and arabinanases.

In aspergilli (**Table 8**), most GH3 xylosidases appear multiple times, averaging three copies per genome, while GH43 xylosidases are represented only once or are absent (*A. niger*).

(iii) **Mannanases and β -mannosidase.** GH2 and GH5 β -mannosidases (EC 3.2.1.25) promote hydrolysis of terminal, nonreducing β -D-mannose residues in β -D-mannosides. GH26 mannan endo- β -1,4-mannosidase (or endo- β -mannanase [EC 3.2.1.78]) promotes random hydrolysis of β -1,4-mannosidic linkages in the main chain of mannans, glucomannans, and galactomannans (166).

The catalytic general acid/base residue is a glutamate, which is separated in sequence by ~100 residues from the other catalytic nucleophile, another glutamate (167). Immediately preceding the general acid/base residue in sequence is an asparagine that makes interactions with the 2-hydroxyl group of the substrate (167).

GH5 and GH26 β -mannanases have been crystallized, and a common three-dimensional structure has been observed. Like other members of these families and xylanases from the GH10 family, they exhibit the typical 8-fold TIM barrel [$(\beta/\alpha)_8$] structure, with the two key active site glutamic acids located at the C-terminal ends of β -strands 4 (acid/base) and 7 (nucleophile) (168–170).

Endo- β -mannanases (GH26) are not common among aspergilli (**Table 8**): only four species (*A. flavus*, *A. nidulans*, *A. niger*, and *A. oryzae*) present this type of endo-acting mannanase, while other mannosidases (end-cutting or exo types) are present more frequently and belong to the GH2 and GH5 families (33, 37, 82).

All three types of mannanases have similar three-dimensional structures, suggesting that they function in similar fashions, and the fact that one cuts a mannan chain randomly (endo) and another splits off dimers from a polymeric substrate is more related to the spatial location of the catalytic amino acids within the substrate-binding domain than to the three-dimensional structure itself (168–170).

Side chain-hydrolyzing enzymes. Carbohydrate side chain-hydrolyzing enzymes are enzymes that hydrolyze sugars linked to the main chain of hemicellulose. Arabinose is the second most abundant sugar in hemicellulose and pectin (171), being found in arabinoxylan and arabinan. Arabinoxylan is constituted by a β -1,4-linked xylopyranose backbone with heterogeneous side chains, such as L-arabinose, O-acetyl, ferulic acid, *p*-coumaric acid, and 4-O-methylglucuronic acid (172). Arabinan is formed by an α -1,2-, α -1,3-, and α -1,5-L-arabinofuranosidic bonds in L-arabinose-containing hemicelluloses, such as arabinoxylan and L-arabinan (174, 175).

We found three families of arabinofuranosidases in fungi: GH51, GH54, and GH62 (**Table 9**). Some operate with wide sub-

TABLE 8 Fungal hemicellulases^a

Fungus	Presence of CBM1				β-Xylanase(s) (EC 3.2.1.8)				β-Xylosidase(s) (EC3.2.1.37)				No. of xylanase/xylosidase gene models per genome				No. of mannosidase gene models per genome			
	GH10	GH11	ACLA_048770, ACLA_086910	ACLA_063140, ACLA_085410, ACLA_062400	ACLA_018590, ACLA_058100	ACLA_018590, ACLA_058100	8	9	ACLA_066240, ACLA_083570	ACLA_066420	3	4	GH2	GH5	GH26	± CBM	Total	± CBM	Total	
<i>A. fumigatus</i> NRRL3357	No	AFLA_008110, AFLA_065510, AFLA_066150	AFLA_090240, AFLA_055190, AFLA_138360, AFLA_010870	AFLA_082390, AFLA_011080, AFLA_039570	AFLA_079500	11	12	AFLA_117830, AFLA_116950, AFLA_128610	AFLA_116950	AFLA_038730	5	6	AFLA_044470	AFLA_044470	AFLA_069870	1	1	4		
<i>A. fumigatus</i> A293	Yes	AFLA_009060				1														
<i>A. nidulans</i> A4	No	AFUA_4G09480, AFUA_3G15210	AFUA_3G00320, AFUA_3G00470, AFUA_3G02090	AFUA_1G16920, AFUA_8G04710	AFUA_8G04710	8	9	AFUA_7G01320, AFUA_6G08840	AFUA_7G01070	AFUA_7G01070	3	4	AFUA_8G07030	AFUA_8G07030	AFUA_8G07030	1	1	4		
<i>A. nidulans</i> A4	Yes	AFUA_6G13610				1														
<i>A. niger</i> CBS513.88	No	ANJ1818, AN2356, AN7401	AN9365, AN3613	AN2359, AN8401, AN2217	AN1477	9	9	AN3368, AN1742 AN9276	AN3358, AN7639, AN2709, AN3297, AN6427	AN7413, AN3336, AN3226	10	11				1	1	2		
<i>A. niger</i> CBS513.88	Yes					0														
<i>A. oryzae</i> RIB40	No	ANL_1_106034	ANL_1_94014, ANL_1_1008124, ANL_1_1968014, ANL_1_1486134	ANL_1_1358014, ANL_1_380154	ANL_1_1358014, ANL_1_2176094	7	7	ANL_1_250104, ANL_1_2176094	ANL_1_160044	ANL_1_1806134	4	4				0	0	0	0	0
<i>A. oryzae</i> RIB40	Yes					0														
<i>A. terreus</i> NIH2624	No	AOR_1_694034, AOR_1_352164, AOR_1_1556114	AOR_1_38094, AOR_1_174164, AOR_1_198014, AOR_1_250034	AOR_1_244054, AOR_1_1726174, AOR_1_212034	AOR_1_1220174	11	12	AOR_1_372024, AOR_1_1298174, AOR_1_1472164	AOR_1_1352024, AOR_1_742074	AOR_1_90054	6	6				0	0	0	0	0
<i>A. terreus</i> NIH2624	Yes	AOR_1_528034				1														
<i>A. terreus</i> NIH2624	No	ATEG_08906, ATEG_00809, ATEG_07190	ATEG_07461, ATEG_04943, ATEG_07383	ATEG_05106, ATEG_09052, ATEG_07383	ATEG_00093	9	10	ATEG_08684, ATEG_09890, ATEG_06636	ATEG_08654, ATEG_02669, ATEG_09991	ATEG_08654, ATEG_02669, ATEG_09991	6	6				0	0	0	0	0
<i>A. terreus</i> NIH2624	Yes	ATEG_03410				1														
Total no. of gene models		22	22	18	6	68	17	18	6	41										
Other																				
<i>Hypocreaficiformis</i>	No	BAA89465/xyn3	XYN2-TRIRE, XYNL-TRIRE	CAA93248	4	0	4	EGR50872, EGR48374	AAA34208	1	3									
<i>N. crassa</i> OR74A	No	NCU05924, NCU08189, NCU07225	NCU02855, NCU0709	NCU0923, NCU01900	8	9	NCU00890	NCU11068		2	2									
<i>N. crassa</i> OR74A	Yes	NCU04997				1	1				0									
Total no. of gene models		5	4	3	1	13	13	3	1	1	5	5	5	5	5	5	5	5	5	5

^a ±, with or without.

TABLE 9 Fungal accessory enzymes

Fungus	Presence of CBM1		Arabinofuranosidase(s)		Arabinanase(s)		Xylan α -1,2-glucuronosidase(s)		Xyloglucanase(s)		No. of accessory genes per genome
	GH51	GH54	GH62	GH43	GH93	GH67	GH15	GH12	GH74	Total	
<i>Aspergillus clavatus</i> NRRL1	No	ACLA_099110, ACLA_074120	ACLA_066470 ACLA_071560	ACLA_042100, ACLA_098980, ACLA_037080, ACLA_018180, ACLA_073460	ACLA_006360	ACLA_029940	AFLA_086510, AFLA_066760, AFLA_116120	AFLA_029160, AFLA_039160	AFLA_029160, AFLA_039160	12	14
<i>A. flavus</i> NRRL3357	No	AFLA_089770, AFLA_015670, AFLA_097160	AFLA_104300	AFLA_063490	AFLA_059500, AFLA_123690, AFLA_085590, AFLA_105970	AFLA_028000, AFLA_039580	AFLA_028000, AFLA_123690, AFLA_085590, AFLA_105970	AFLA_017270	AFLA_044310	2	
<i>A. fumigatus</i> Af293	No	AFUA_2G15160	AFUA_6G14620	AFUA_2G00920 AFUA_2G12770	AFUA_3G14620, AFUA_6G00770, AFUA_1G17320, AFUA_2G14150	AFUA_2G04570, AFUA_6G12120	AFUA_7G04680	AFUA_1G04730	AFUA_8G01490	13	14
<i>A. nidulans</i> A4	No	AN2541, AN9520	AN1571	AN2632, AN7908	AN2534, AN8007, AN6352, AN3044	AN2060	AN9329	AN0452	AN1542	13	15
<i>A. niger</i> CBS513.88	No	AN1_1_42014, AN1_1_1010084	AN1_1_372134	AN1_1_108034	AN1_1_118084, AN1_1_1480024, AN1_1_2268024, AN1_1_390144	AN1_1_1814154, AO1_1_246054	AN1_1_1814154, AO1_1_246054	AN1_1_432014	AN5061	2	11
<i>A. oryzae</i> RB40	No	AOR_1_46104, AOR_1_51894, AOR_1_124084	AOR_1_144	AOR_1_1552114, AO090103000088	AOR_1_882114, AO1_1_90064, AO1_1_2296174, AO1_1_1872144	AOR_1_1814154, AO1_1_246054	AOR_1_458164, AO1_1_3226174, AO1_1_54024, AO1_1_1370094	AN1_1_236017	2		
<i>A. terreus</i> NIH2624	No	ATEG_02882, ATEG_03540, ATEG_07868	ATEG_07939	ATEG_10071, ATEG_10379, ATEG_00186	ATEG_03520, ATEG_03688, ATEG_01407, ATEG_01562	ATEG_06045, ATEG_00891	ATEG_09975	ATEG_0355, ATEG_09974	ATEG_03755	17	19
Total no. of gene models	16	7	13	29	9	8	12	8	6	108	
Other <i>Hypocreales</i>	No	ABFL_HYPIE, EGR51797	AAP57750	EGR47214	4	6			XG74_TRIRE	2	
<i>N. crassa</i> OR74A	No	NCU02343	NCU09775	NCU09924	NCU06143	NCU07351	NCU05955	NCU05955	4	6	
Total no. of gene models	1	3	1	1	2	2	2	2	2	12	12

^a, with or without.

strate specificity, acting on arabinofuranosides at O-5, O-2, and/or O-3 as a single substituent (171). GH51 arabinofuranosidases hydrolyze small substrates only (176), GH54 enzymes hydrolyze polymeric substrates, such as arabinoxylans, in addition to the small substrates, and GH62 enzymes act only on arabinoxylans (151, 176).

The three-dimensional structures of the catalytic domains of fungal GH51 and GH62 arabinofuranosidases are not yet known. However, the bacterial enzymes show a classical $(\beta/\alpha)_8$ barrel structure, and there is also evidence that arabinofuranosidases act as hexamers (177, 178). The three-dimensional structure of GH54 arabinofuranosidase establishes a β -sandwich (176).

Table 9 shows that genes encoding arabinofuranosidases are abundant in aspergilli, averaging three GH51, one GH54, and two GH62 gene models per genome. None of the examined arabinofuranosidases are linked to a CBM.

(ii) Arabinanases. Most monosaccharides are present in their D-form. L-Arabinose is the exception and is found in its furanose form as a constituent of hemicellulose and pectin (171). Arabinanases and arabinofuranosidases catalyze the hydrolysis of α -1,5-arabinofuranosidic bonds in arabinose-containing polysaccharides (151, 179). GH43 and GH93 arabinanases share extensive amino acid sequence similarities with β -fructosidases (GH32 and GH68), with multiple homologous domains (180).

Arabinanases act by depolymerizing arabinopolysaccharides, producing arabinose or arabino oligomers, depending on the preference of the enzyme for substrate termini (exo type) (181) or random cleavage (endo type) (182).

Three-dimensional structures for fungal GH43 arabinanases are not yet available, and the closest crystallized arabinanase is from *Cellvibrio japonicus* (E value = 2E-35) and displays the classical five-blade β -propeller fold (181, 183). A V-shaped groove that is partially enclosed at one end forms a single extended substrate-binding surface across the face of the propeller (183), and three carboxylates deep in the active site provide the general acid and base components for glycosidic bond hydrolysis, with inversion of the anomeric configuration (183).

GH93 arabinanases recognize linear α -1,5-L-arabinan as the substrate and release α -1,5-L-arabinobiose from the nonreducing end of the polysaccharide (81). Two fungal GH93 arabinanases (81, 184) show a six-blade β -propeller fold with a typical "Velcro ring" closure. The substrate-binding groove is enclosed at one end by two residues, Glu and Tyr, which contribute to the recognition of the nonreducing chain end of the polysaccharide (184).

Table 9 shows that genes encoding arabinanases are also abundant, with the genomes averaging five GH43 and one GH93 arabinanase gene. None of the arabinanases are linked to CBMs.

(iii) β -Glucuronidases and xylan α -1,2-glucuronosidase. Glucuronidases cleave the glucuronic acid of glucuronoglycans found in plant cell wall structures. Two types of glucuronoglycans are common: glucuronans (alginic acid), constituted exclusively of glucuronic acid residues, and glucuronoglycans (pectins, gums, and mucilages), whose main chain is composed of glucuronic acid and other sugars as side chains (185). Glucuronans are often acetylated, which interferes with glucuronan-hydrolyzing enzymes (186).

Aspergillus glucuronidases are found in two CAZy families, GH2 and GH79, and none of them have been characterized. **Table 9** shows that β -glucuronidases are much less abundant than arabinofuranosidases and arabinanases, averaging one or no GH2 or

GH79 gene model. Xylan α -1,2-glucuronidases are found in two CAZy families, GH67 and GH115, and none have been characterized. **Table 9** shows that xylan α -1,2-glucuronosidases are represented at averages of 1 and 2 genes for GH67 and GH115 enzymes, respectively.

(iv) Xyloglucanases. Xyloglucan consists of a cellulose-like backbone chain of β -1,4-glucan with xylosyl side chains attached to the O-6 position of glycosyl residues and associates with cellulose microfibrils through hydrogen bonds, forming cellulose-xyloglucan (187). Enzymes responsible for the hydrolysis of the xyloglucan backbone are xyloglucan endo- β -1,4-glucanases or "xyloglucanases" (EC 3.2.1.151). EC 3.2.1.151 reflects many different enzyme sequences, structures, and hydrolytic mechanisms, with either inversion or retention of the anomeric carbon. In the sequence-based CAZy classification, enzymes defined as xyloglucanases are found in retaining families GH5, GH12, and GH16 and inverting families GH44 and GH74 (188). To date, aspergillus GH12 and GH74 xyloglucanases are the best described (189).

While some endoglucanases are promiscuous and can hydrolyze both unbranched and branched β -1,4-glucan chains (189), xyloglucan-specific endoglucanases (XEGs) constitute a relatively new class of enzymes (190) and were described as EC 3.2.1.151 and EC 3.2.1.150 enzymes (191).

In EC 3.2.1.151 XEGs, the reaction involves endo hydrolysis of β -1,4-D-glycosidic linkages in xyloglucan, with retention of the beta-configuration of the glycosyl residues (192–194). In EC 3.2.1.150 XEGs (oligoxyloglucan reducing-end-specific cellobiohydrolases [OREX]), the reaction involves the hydrolysis of cellobiose from the reducing end of xyloglucans consisting of a β -1,4-linked glucan carrying alpha-D-xylosyl groups on O-6 of the glucose residues (195).

OREX belong to the GH74 family. Two GH74 enzymes have been shown to be reducing-end-specific cellobiohydrolases (EC 3.2.1.150) (195, 196) releasing XG, LG, or FG from xyloglucan [X, G, L, and F designate α -D-Xylp-(1 \rightarrow 6)- β -D-GlcP, D-GlcP, β -D-Galp-(1 \rightarrow 2)- α -D-Xylp-(1 \rightarrow 6)- β -D-GlcP, and α -L-Fucp-(1 \rightarrow 2)- β -D-Galp-(1 \rightarrow 2)- α -D-Xylp-(1 \rightarrow 6)- β -D-GlcP, respectively] (197). The exo-acting activity is believed to be due to a loop insertion closing off the positive subsites (198).

The OREX X-ray crystal structure shows two seven-blade β -propeller domains forming a large cleft and a loop where the substrate binds (Fig. 5). The substrate-cleaving region is located near the loop region, believed to be the region of hydrolysis. Deletion of the loop region resulted in endo random cleavages of the substrate, suggesting that the loop region directs the exo activity (198–200). Moreover, an endo-processive xyloglucanase (XEG74) that contains four unique tryptophan residues, in the negative subsites (W61 and W64) and the positive subsites (W318 and W319), was isolated from *Paenibacillus*. The positive subsites (W318 and W319) were essential for processive degradation and were responsible for maintaining binding interactions with xyloglucan (201).

Xyloglucanases are not abundant in fungi. According to CAZy and mycoCLAP (202), to date, there are no GH5, GH16, or GH44 XEGs characterized from fungi. Thus, we considered only GH12 and GH74 XEGs from aspergilli, due to the lack of characterized gene models for the GH5, GH16, and GH44 families. Furthermore, XEGs from these families were not identified in secretomes of aspergilli growing on biomass (10, 14, 203, 204). **Table 9** shows that aspergilli contain one or no GH74 xyloglucanases but several GH12 enzymes. *A. flavus* and *A. oryzae* have no genes encoding

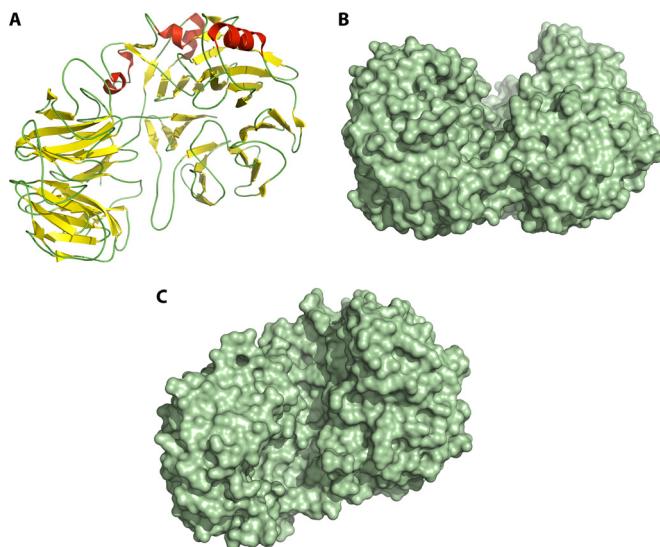


FIG 5 GH74 xyloglucanobiohydrolases. (A) The three-dimensional structure consists of two tandem repeats of a seven-blade β -propeller domain which forms a large cleft and a loop where the substrate is bound. (B and C) Two views of the open cleft. The three-dimensional structure of *Geotrichum* sp. GH74 endoglucanase is from PDB entry 1SQJ (199, 200).

GH74 xyloglucanases, and *A. nidulans* has two. Other aspergilli have one GH74 xyloglucanase. Unlike *A. flavus*, all other aspergilli contain one GH12 xyloglucanase. GH12 xyloglucanases are not linked to a CBM (see Table S7 in the supplemental material).

Accessory enzymes. (i) **Acetyl xylan esterases.** Acetyl xylan esterases promote hydrolysis of acetyl groups from xylan β -1,4-linked D-xylopyranoside backbones with heterologous side chains (O-acetyl, ferulic acid, p-coumaric acid, arabinose, and 4-O-methylglucuronic acid groups), mainly present in plant cell walls of hardwoods.

Few acetyl xylan esterases have been studied so far, with one from *A. niger* belonging to carbohydrate esterase family 1 (CE1) and two CE5 enzymes, from *T. reesei* and *Penicillium purpurogenum*, being described (205–207). *A. awamori* CE1 acetyl xylan esterase (accession no. XP_001395572) releases acetic acid from birchwood and is active toward α -naphthylacetate (C_2) and α -naphthylpropionate (C_3) but not toward acyl substrates containing four or more carbons (206), and site-directed mutagenesis that abolished deacetylation indicated that Ser₁₁₉ and Asp₂₀₂ are key. *T. reesei* CE5 acetyl xylan esterase (AXE1_TRIRE) catalyzes acetylated xylo-oligomers and releases acetic acid from birchwood xylan (207). A novel acetyl xylan esterase (Aes1) from *T. reesei*, with no CAZy CE classification, was studied using 2-, 3-, and 4-O-acetyl 4-nitrophenyl D-xylopyranosides as substrates, and Aes1 hydrolyzed the three substrates with an initial rate ratio of 1:19:17.7, suggesting that Aes1 prefers positions 3 and 4, while the other acetyl xylan esterases prefer position 2 acetyl groups (208–210).

Carbohydrate esterases are classified into 16 CAZy CE families, whereas fungal acetyl xylan esterases appear in 8 of them (211). Most carbohydrate esterases are serine-type esterases also acting on low-molecular-weight substrates, such as 4-nitrophenyl or 4-methylumbelliferyl acetate. An exception is the acetyl xylan esterases from family CE4, which do not operate on aryl acetates (212, 213).

The 3-D structure of acetyl xylan esterase consists of a three-layer $\alpha/\beta/\alpha$ sandwich fold (214). The central β -sheet consists of six parallel β -strands delimited by four α -helices (two on each side). In addition, acetyl xylan esterase docks an additional two α -helices and four β_{10} -helices (214). In acetyl xylan esterases, cysteine forms disulfide bridges, and the active site is located in a cleft near the C-terminal end of the third β -strand, where there is a Ser-His-Asp triad (214).

Table 10 shows that the better-understood acetyl xylan esterases belong to the CE1 family; however, aspergilli also have acetyl xylan esterases belonging to the CE3, CE4, CE5, and CE16 families. Most aspergilli harbor genes belonging to at least three or four families in their genomes, with the exception of *A. oryzae*, which has only two CE families represented. Acetyl xylan esterases belonging to families CE1 and CE5 are sometimes linked to a CBM (Table 10).

(ii) **Ferulic acid esterases.** Ferulic acid (4-hydroxy-3-methoxy-cinnamic acid), a hydroxycinnamic acid, is the main aromatic acid building block of lignocellulosic materials (215, 216). Ferulic acid is generally not found free but instead is esterified to arabinose in various polysaccharides, e.g., arabinoxylans and pectins (217).

Feruloyl esterase, also known as cinnamoyl esterase, is an enzyme that hydrolyzes ferulate ester groups from hemicellulose molecules, which represent cross-links among hemicellulose chains and lignin polymers (218). Feruloyl esterases belong to the CE1 family and hydrolyze the ester bond between hydroxycinnamic acid and sugars present in plant cell walls (219).

The three-dimensional structure of *A. niger* feruloyl esterase (PDB entries 1UWC and 1USW) displays an α/β hydrolase fold consisting of a nine-stranded β -sheet core surrounded by α -helices and two additional β -strands (220). The catalytic triad (Ser133-His247-Asp194) forms the active site of this enzyme. The active site cavity is confined by a lid, analogous to the case in lipases, and by a loop that confers plasticity to the substrate-binding site (221). Feruloyl esterases were initially classified by aromatic functional categories (222, 223), and later the classification was extended to subfamilies A, B, C, and D, based on primary amino acid sequence similarity and substrate specificity against four model substrates, i.e., methyl 3-methoxy-4-hydroxycinnamate (MFA), methyl 3,4-dihydroxycinnamate (MCA), methyl 4-hydroxycinnamate (MpCA), and methyl 3,5-dimethoxy-4-hydroxycinnamate (MSA) (224, 225).

Table 10 shows that feruloyl esterases are represented in variable copy numbers in all aspergilli, in some cases showing much more than 2 copies (3 copies in *A. clavatus*, 8 in *A. flavus* and *A. terreus*, 10 in *A. oryzae*, and 11 copies in *A. niger*). None of the feruloyl esterases are linked to CBMs.

Aspergillus hemicellulase degradation summary. Aspergilli contain an average of 30 hemicellulase gene models (Tables 8 to 10), while *H. jecorina* and *N. crassa* contain an average of 17 each, distributed among the following seven distinct functional categories: the backbone-attacking enzymes xylanase, mannosidase, arabinase, and xyloglucanase and the short-side-chain-removing enzymes xylan α -1,2-glucuronidase, arabinofuranosidase, and xylosidase. In addition, aspergilli contain, on average, seven accessory gene models involved in the liberation of acetyl groups from acetylated polysaccharides and in cleavage of the ester links between monomeric or dimeric ferulic acid and the polysaccharide main chain of xylans.

TABLE 10 Genome-wide acetyl xyran and feruloyl esterase content in aspergilli

Fungus	Acetyl xyran esterase(s) (EC 3.1.1.72)				No. of acetyl xyran esterase gene models per genome				No. of feruloyl esterase gene models in genome			
	Other domain	CE1	CE3	CE5	CE16	± domain	Total	CE1 feruloyl esterase(s) (EC 3.1.1.73)	± domain	Total		
<i>A. fumigatus</i>	None	ACLA_081220										
<i>A. clavatus</i> NRRL1	CBM1											
<i>A. flavus</i> NRRL3357	None	AFLA_045570										
XynB												
<i>A. fumigatus</i> AF293	None	AFLA_073610, AFLA_063680										
CBM1	AFU_A_8G06570	AFUA_1G03170	AFUA_2G00630, AFUA_4G14700	3	5	AFUA_2G14530, AFUA_5G00960, AFUA_6G00050, AFUA_6G00940	4	AFUA_105900, AFLA_120560, AFLA_110270, AFLA_135510, AFLA_063480, AFLA_000910	8	0	3	3
<i>A. nidulans</i> A4	None	AN6093	AN6464, AN9260	AN1792	2	4	AN5267.2, AN1772.2	2	0	2	2	
XynB												
<i>A. niger</i> CBS513.88	None	ANL_1_652104										
CBM1												
<i>A. oryzae</i> RIB40	None	AOR_1_1260054	AOR_1_576114	2	AOR_1_1866144, AOR_1_1042164, AOR_1_1060144, AOR_1_350164, AOR_1_20134, AOR_1_112124, AOR_1_102164, AOR_1_958024, AOR_1_412034, AOR_1_155014	10						
XynB			ANL_1_1582114		1							
<i>A. terreus</i> NIH2624	None	ATEG_09843	ATHG_04709, ATEG_04056	ATEG_08115	4	ATEG_08112, ATEG_02212, ATEG_09113, ATEG_06644, ATEG_06663, ATEG_01914, ATEG_08907, ATEG_02415	8					
XynB			ATEG_09958	ATEG_08874	2	0						
Total no. of gene models	7	7	4	8	26	46	46					
Other												
<i>Hypocreaficiformis</i>	None	EGR53049	AB134466, EGR51758	3	5	0	0					
XynB	CBM1	EGR44597	CAA93247	1	1	0	0					
<i>N. crassa</i> OR74A	None	NCU04870, NCU04494	NCU09664	NCU06364	4	6	NCU09774, NCU04870, NCU0926, NCU09491, NCU08785	5	0	5	5	
Total no. of gene models	3	2	3	3	11	11	5	5	5	5	5	

CONCLUSIONS

Industrial enzymology is a billion-dollar market that capitalizes on the inherent catalytic specificity, speed, and robustness of natural enzymes to carry out complex reactions in a clean and environmentally friendly manner. Holocellulose-degrading enzyme applications are well established within the paper, food, and feedstock industries. Cellulase and hemicellulase cocktails are essential components in any kind of biorefinery core that relies on biomass materials as the input source.

In the present study, based on the genome-wide content of seven aspergilli (a natural biomass biorefinery), we unraveled hundreds of gene models encoding holocellulose-degrading enzymes, suggesting the occurrence of dozens of apparent duplications, but after a systematic organization, these fell into smaller coherent functional groups according to domain organization, biochemical activity, and genome distribution.

Current commercial enzyme cocktails are catalytically incomplete, demanding high enzyme loads and long residence times, and are subject to contamination. Most commercially available cocktails lack one or more enzyme activities that we found in the genome-wide survey presented here.

The present study provides a valuable reference on carbohydrate degradation by aspergilli, giving a significant extension of previous studies (33, 34, 37). The mapping of holocellulose enzyme breakdown systems in filamentous fungi and their relationships with substrates provides the foundation for genome-wide expression analysis studies, as well as a powerful framework for further functional and structural works on fungal carbohydrate-active enzymes.

ACKNOWLEDGMENTS

We thank the many labs that invested in biomass biotechnology as well as all the microbial researchers who discovered amazing polymer breakdown enzymes. We regret not being able to cite all of the work that has shaped these fields over the last several decades. We also thank anonymous referees for their insightful comments and valuable suggestions, which improved the manuscript.

Research done by us and cited in this report received support from the Oklahoma Bioenergy Center, Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant 310177/2011-1 to F.M.S.), Fundação de Amparo à Pesquisa do Estado de São Paulo (grants 2012/20549-4 and 2013/18910-3 to A.R.L.D. and grants 2008/58037-9 and 2010/18198-3 to F.M.S.), the National Renewable Energy Laboratory (grant ZDJ-7-77608-01 to R.A.P.), the U.S. Department of Agriculture (CSREES grant 2007-35504-18244 to R.A.P.), and the U.S. Department of Energy and Edenspace Corp. (grant 06103-OKL to R.A.P.).

REFERENCES

- Hawkins AS, McTernan PM, Lian H, Kelly RM, Adams MW. 2013. Biological conversion of carbon dioxide and hydrogen into liquid fuels and industrial chemicals. *Curr. Opin. Biotechnol.* 24:376–384. <http://dx.doi.org/10.1016/j.copbio.2013.02.017>.
- Amorim HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH. 2011. Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 91:1267–1275. <http://dx.doi.org/10.1007/s00253-011-3437-6>.
- Lal R. 2005. World crop residues production and implications of its use as a biofuel. *Environ. Int.* 31:575–584. <http://dx.doi.org/10.1016/j.envint.2004.09.005>.
- Wei H, Xu Q, Taylor LE, 2nd, Baker JO, Tucker MP, Ding SY. 2009. Natural paradigms of plant cell wall degradation. *Curr. Opin. Biotechnol.* 20:330–338. <http://dx.doi.org/10.1016/j.copbio.2009.05.008>.
- Chassard C, Delmas E, Robert C, Bernalier-Donadille A. 2010. The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol. Ecol.* 74:205–213. <http://dx.doi.org/10.1111/j.1574-6941.2010.00941.x>.
- Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331:463–467. <http://dx.doi.org/10.1126/science.1200387>.
- Rattanachomsri U, Kanokratana P, Eurwilaichitr L, Igarashi Y, Champreda V. 2011. Culture-independent phylogenetic analysis of the microbial community in industrial sugarcane bagasse feedstock piles. *Biosci. Biotechnol. Biochem.* 75:232–239. <http://dx.doi.org/10.1271/bbb.100429>.
- Ligganstoffer AS, Youssef NH, Couger MB, Elshahed MS. 2010. Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigmota) in ruminant and non-ruminant herbivores. *ISME J.* 4:1225–1235. <http://dx.doi.org/10.1038/ismej.2010.49>.
- Liu N, Yan X, Zhang M, Xie L, Wang Q, Huang Y, Zhou X, Wang S, Zhou Z. 2011. Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. *Appl. Environ. Microbiol.* 77:48–56. <http://dx.doi.org/10.1128/AEM.01521-10>.
- Saykhedkar S, Ray A, Ayoubi-Canaan P, Hartson SD, Prade R, Mort AJ. 2012. A time course analysis of the extracellular proteome of *Aspergillus nidulans* growing on sorghum stover. *Biotechnol. Biofuels* 5:52. <http://dx.doi.org/10.1186/1754-6834-5-52>.
- Berka RM, Grigoriev IV, Ottillar R, Salamov A, Grimwood J, Reid I, Ishmael N, John T, Darmond C, Moisan MC, Henrissat B, Coutinho PM, Lombard V, Natvig DO, Lindquist E, Schmutz J, Lucas S, Harris P, Polowolski J, Bellemare A, Taylor D, Butler G, de Vries RP, Allijin IE, van den Brink J, Ushinsky S, Storms R, Powell AJ, Paulsen IT, Elbourne LD, Baker SE, Magnuson J, Laboissiere S, Clutterbuck AJ, Martinez D, Wogulis M, de Leon AL, Rey MW, Tsang A. 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat. Biotechnol.* 29:922–927. <http://dx.doi.org/10.1038/nbt.1976>.
- van den Brink J, van Muiswinkel GC, Theelen B, Hinz SW, de Vries RP. 2013. Efficient plant biomass degradation by thermophilic fungus *Myceliophthora heterothallica*. *Appl. Environ. Microbiol.* 79:1316–1324. <http://dx.doi.org/10.1128/AEM.02865-12>.
- Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. 2012. Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* 5:45. <http://dx.doi.org/10.1186/1754-6834-5-45>.
- Ray A, Saykhedkar S, Ayoubi-Canaan P, Hartson SD, Prade R, Mort AJ. 2012. *Phanerochaete chrysosporium* produces a diverse array of extracellular enzymes when grown on sorghum. *Appl. Microbiol. Biotechnol.* 93:2075–2089. <http://dx.doi.org/10.1007/s00253-012-3907-5>.
- Phillips CM, Beeson WT, Cate JH, Marletta MA. 2011. Cellulose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem. Biol.* 6:1399–1406. <http://dx.doi.org/10.1021/cb200351y>.
- Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, Cate JH, Glass NL. 2012. Induction of lignocellulose-degrading enzymes in *Neurospora crassa* by cellobextrins. *Proc. Natl. Acad. Sci. U. S. A.* 109:6012–6017. <http://dx.doi.org/10.1073/pnas.1118440109>.
- Carvalheiro F, Duarte LC, Gírio FM. 2008. Hemicellulose biorefineries: a review on biomass pretreatments. *J. Sci. Ind. Res.* 67:849–864.
- Chundawat SP, Beckham GT, Himmel ME, Dale BE. 2011. Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2:121–145. <http://dx.doi.org/10.1146/annurev-chembioeng-061010-114205>.
- Ferrer M, Martínez-Abarca F, Golyshev PN. 2005. Mining genomes and ‘metagenomes’ for novel catalysts. *Curr. Opin. Biotechnol.* 16:588–593. <http://dx.doi.org/10.1016/j.copbio.2005.09.001>.
- Himmel ME, Bayer EA. 2009. Lignocellulose conversion to biofuels: current challenges, global perspectives. *Curr. Opin. Biotechnol.* 20:316–317. <http://dx.doi.org/10.1016/j.copbio.2009.05.005>.
- Otero JM, Panagiotou G, Olsson L. 2007. Fueling industrial biotechnology growth with bioethanol. *Adv. Biochem. Eng. Biotechnol.* 108:1–40.
- Regalbuto JR. 2009. Engineering Cellulosic biofuels—got gasoline? *Science* 325:822–824. <http://dx.doi.org/10.1126/science.1174581>.
- Regalbuto JR. 2011. The sea change in US biofuels’ funding: from cellulosic ethanol to green gasoline. *Biofuels Bioprod. Biorefining* 5:495–504. <http://dx.doi.org/10.1002/bbb.298>.
- Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K, Arima T, Akita O, Kashiwagi Y, Abe K, Gomi K, Horiuchi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE,

- Nierman WC, Yu J, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoyama A, Ichinomiya M, Igarashi R, Iwashita K, Juvvadi PR, Kato M, Kato Y, Kin T, Kokubun A, Maeda H, Maeyama N, Maruyama J, Nagasaki H, Nakajima T, Oda K, Okada K, Paulsen I, Sakamoto K, Sawano T, Takahashi M, Takase K, Terabayashi Y, Wortman JR, Yamada O, Yamagata Y, Anazawa H, Hata Y, Koide Y, Komori T, Koyama Y, Minetoki T, Suharnan S, Tanaka A, Isono K, Kuhara S, Ogasawara N, Kikuchi H. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 438:1157–1161. <http://dx.doi.org/10.1038/nature04300>.
25. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berrianan M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulson R, Davies R, Dyer PS, Farman M, Fedorova N, Fedorova N, Feldblum TV, Fischer R, Fosker N, Fraser A, Garcia JL, Garcia MJ, Goble A, Goldman GH, Gomi K, Griffith-Jones S, Gwilliam R, Haas B, Haas H, Harris D, Horiuchi H, Huang J, Humphray S, Jimenez J, Keller N, Khouri H, Kitamoto K, Kobayashi T, Konzack S, Kulkarni R, Kumagai T, Lafon A, Latge JP, Li W, Lord A, Lu C, Majoros WH, May GS, Miller BL, Mohamoud Y, Molina M, Monod M, Mouyna I, Mulligan S, Murphy L, O’Neil S, Paulsen I, Penalva MA, Pertea M, Price C, Pritchard BL, Quail MA, Rabinowitz E, Rawlins N, Rajandream MA, Reichard U, Renaud H, Robson GD, Rodriguez de Cordoba S, Rodriguez-Pena JM, Ronning CM, Rutter S, Salzberg SL, Sanchez M, Sanchez-Ferrero JC, Saunders D, Seeger K, Squares R, Squares S, Takeuchi M, Tekaia F, Turner G, Vazquez de Aldana CR, Weidman J, White O, Woodward J, Yu JH, Fraser C, Galagan JE, Asai K, Machida M, Hall N, Barrell B, Denning DW. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438:1151–1156. <http://dx.doi.org/10.1038/nature04322>.
26. Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JA, van den Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutinho PM, Danchin EG, Debets AJ, Dekker P, van Dijken PW, van Dijken A, Dijkhuizen L, Driessens AJ, d’Enfert C, Geysens S, Goosen C, Groot GS, de Groot PW, Guillemette T, Henrissat B, Herweijer M, van den Hombergh JP, van den Hondel CA, van der Heijden RT, van der Kaaij RM, Klis FM, Kools HJ, Kubicek CP, van Kuyk PA, Lauber J, Lu X, van der Maarel MJ, Meulenbergh R, Menke H, Mortimer MA, Nielsen J, Oliver SG, Olsthoorn M, Pal K, van Peij NN, Ram AF, Rinas U, Roubos JA, Sagt CM, Schmoll M, Sun J, Ussery D, Varga J, Vervecken W, van de Vondervoort PJ, Wedler H, Wosten HA, Zeng AP, van Ooyen AJ, Visser J, Stam H. 2007. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.* 25:221–231. <http://dx.doi.org/10.1038/nbt1282>.
27. Payne GA, Nierman WC, Wortman JR, Pritchard BL, Brown D, Dean RA, Bhatnagar D, Cleveland TE, Machida M, Yu J. 2006. Whole genome comparison of *Aspergillus flavus* and *A. oryzae*. *Med. Mycol.* 44:S9–S11. <http://dx.doi.org/10.1080/13693780600835716>.
28. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D’Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Penalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paolelli M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438:1105–1115. <http://dx.doi.org/10.1038/nature04341>.
29. Arnaud MB, Chibucus MC, Costanzo MC, Crabtree J, Inglis DO, Lotia A, Orvis J, Shah P, Skrzypek MS, Binkley G, Miyasato SR, Wortman JR, Sherlock G. 2010. The *Aspergillus* Genome Database, a curated comparative genomics resource for gene, protein and sequence information for the *Aspergillus* research community. *Nucleic Acids Res.* 38:D420–D427. <http://dx.doi.org/10.1093/nar/gkp751>.
30. Minic Z, Joulian L. 2006. Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol. Biochem.* 44:435–449. <http://dx.doi.org/10.1016/j.plaphy.2006.08.001>.
31. Ward OP, Moo-Young M. 1989. Enzymatic degradation of cell wall and related plant polysaccharides. *Crit. Rev. Biotechnol.* 8:237–274. <http://dx.doi.org/10.3109/07388558909148194>.
32. Culleton H, McKie V, de Vries RP. 2013. Physiological and molecular aspects of degradation of plant polysaccharides by fungi: what have we learned from *Aspergillus*? *Biotechnol. J.* 8:884–894. <http://dx.doi.org/10.1002/biot.201200382>.
33. van den Brink J, de Vries RP. 2011. Fungal enzyme sets for plant polysaccharide degradation. *Appl. Microbiol. Biotechnol.* 91:1477–1492. <http://dx.doi.org/10.1007/s00253-011-3473-2>.
34. Andersen MR, Giese M, de Vries RP, Nielsen J. 2012. Mapping the polysaccharide degradation potential of *Aspergillus niger*. *BMC Genomics* 13:313. <http://dx.doi.org/10.1186/1471-2164-13-313>.
35. Coutinho PM, Andersen MR, Kolenova K, vanKuyk PA, Benoit I, Gruben BS, Trejo-Aguilar B, Visser H, van Solingen P, Pakula T, Seibold B, Battaglia E, Aguilar-Osorio G, de Jong JF, Ohm RA, Aguilar M, Henrissat B, Nielsen J, Stalbrand H, de Vries RP. 2009. Post-genomic insights into the plant polysaccharide degradation potential of *Aspergillus nidulans* and comparison to *Aspergillus niger* and *Aspergillus oryzae*. *Fungal Genet. Biol.* 46(Suppl 1):S161–S169. <http://dx.doi.org/10.1016/j.fgb.2008.07.020>.
36. Martens-Uzunova ES, Schaap PJ. 2009. Assessment of the pectin degrading enzyme network of *Aspergillus niger* by functional genomics. *Fungal Genet. Biol.* 46(Suppl 1):S170–S179. <http://dx.doi.org/10.1016/j.fgb.2008.07.021>.
37. de Vries RP, Visser J. 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol. Mol. Biol. Rev.* 65:497–522. <http://dx.doi.org/10.1128/MMBR.65.4.497-522.2001>.
38. Gray KA, Zhao L, Emptage M. 2006. Bioethanol. *Curr. Opin. Chem. Biol.* 10:141–146. <http://dx.doi.org/10.1016/j.cbpa.2006.02.035>.
39. Kim M, Day DF. 2011. Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. *J. Ind. Microbiol. Biotechnol.* 38:803–807. <http://dx.doi.org/10.1007/s10295-010-0812-8>.
40. Sjöström E. 1993. Wood chemistry. Fundamentals and applications, 2nd ed. Academic Press, San Diego, CA.
41. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315:804–807. <http://dx.doi.org/10.1126/science.1137016>.
42. Hames BR. 2009. Biomass compositional analysis for energy applications. *Methods Mol. Biol.* 581:145–167. http://dx.doi.org/10.1007/978-1-60761-214-8_11.
43. Reddy N, Yang Y. 2005. Biofibers from agricultural byproducts for industrial applications. *Trends Biotechnol.* 23:22–27. <http://dx.doi.org/10.1016/j.tibtech.2004.11.002>.
44. Ding SY, Himmel ME. 2006. The maize primary cell wall microfibril: a new model derived from direct visualization. *J. Agric. Food Chem.* 54:597–606. <http://dx.doi.org/10.1021/jf051851z>.
45. Kovalenko VI. 2010. Crystalline cellulose: structure and hydrogen bonds. *Russ. Chem. Rev.* 79:231. <http://dx.doi.org/10.1070/RC2010v07n03ABEH004065>.
46. Payne CM, Himmel ME, Crowley MF, Beckham GT. 2011. Decrystallization of oligosaccharides from the cellulose I β surface with molecular simulation. *J. Phys. Chem. Lett.* 2:1546–1550. <http://dx.doi.org/10.1021/jz2005122>.
47. Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T, Vorwerk S, Youngs H. 2004. Toward a systems approach to understanding plant cell walls. *Science* 306:2206–2211. <http://dx.doi.org/10.1126/science.1102765>.
48. Vietor RJ, Mazeau K, Lakin M, Perez S. 2000. A priori crystal structure prediction of native celluloses. *Biopolymers* 54:342–354. [http://dx.doi.org/10.1002/1097-0282\(20001015\)54:5<342::AID-BIP50>3.0.CO;2-0](http://dx.doi.org/10.1002/1097-0282(20001015)54:5<342::AID-BIP50>3.0.CO;2-0).
49. Atalla RH, Vanderhart DL. 1984. Native cellulose: a composite of two distinct crystalline forms. *Science* 223:283–285. <http://dx.doi.org/10.1126/science.223.4633.283>.
50. Li Y, Lin M, Davenport JW. 2011. Ab initio studies of cellulose I: crystal structure, intermolecular forces, and interactions with water. *J. Phys. Chem. C* 115:11533–11539. <http://dx.doi.org/10.1021/jp2006759>.
51. Zugemmaier P. 2001. Conformation and packing of various crystalline cellulose fibers. *Prog. Polym. Sci.* 26:1341–1417. [http://dx.doi.org/10.1016/S0079-6700\(01\)00019-3](http://dx.doi.org/10.1016/S0079-6700(01)00019-3).
52. Wada M, Heux L, Sugiyama J. 2004. Polymorphism of cellulose I family: reinvestigation of cellulose IV α . *Biomacromolecules* 5:1385–1391. <http://dx.doi.org/10.1021/bm0345357>.
53. Wada M, Nishiyama Y, Chanzy H, Forsyth T, Langan P. 2008. The

- structure of celluloses. *Powder Diffr.* 23:92–95. doi:<http://dx.doi.org/10.1154/1.2912442>.
54. Hall M, Bansal P, Lee JH, Realff MJ, Bommarius AS. 2010. Cellulose crystallinity—a key predictor of the enzymatic hydrolysis rate. *FEBS J.* 277:1571–1582. <http://dx.doi.org/10.1111/j.1742-4658.2010.07585.x>.
 55. Mittal A, Katahira R, Himmel ME, Johnson DK. 2011. Effects of alkaline or liquid-ammonia treatment on crystalline cellulose: changes in crystalline structure and effects on enzymatic digestibility. *Biotechnol. Biofuels* 4:41. <http://dx.doi.org/10.1186/1754-6834-4-41>.
 56. Wada M, Chanzy H, Nishiyama Y, Langan P. 2004. Cellulose III₁ crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. *Macromolecules* 37:8548–8555. <http://dx.doi.org/10.1021/ma0485585>.
 57. Du Toit PJ, Olivier SP, van Biljon PL. 1984. Sugar cane bagasse as a possible source of fermentable carbohydrates. I. Characterization of bagasse with regard to monosaccharide, hemicellulose, and amino acid composition. *Biotechnol. Bioeng.* 26:1071–1078.
 58. Scheller HV, Ulvskov P. 2010. Hemicelluloses. *Annu. Rev. Plant Biol.* 61: 263–289. <http://dx.doi.org/10.1146/annurev-arplant-042809-112315>.
 59. Gibson LJ. 2012. The hierarchical structure and mechanics of plant materials. *J. R. Soc. Interface*. 9:2749–2766. <http://dx.doi.org/10.1098/rsif.2012.0341>.
 60. Prade RA, Zhan D, Ayoubi P, Mort AJ. 1999. Pectins, pectinases and plant-microbe interactions. *Biotechnol. Genet. Eng. Rev.* 16:361–391. <http://dx.doi.org/10.1080/02648725.1999.10647984>.
 61. Caffall KH, Mohnen D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr. Res.* 344:1879–1900. <http://dx.doi.org/10.1016/j.carres.2009.05.021>.
 62. Davin LB, Lewis NG. 2005. Lignin primary structures and dirigent sites. *Curr. Opin. Biotechnol.* 16:407–415. <http://dx.doi.org/10.1016/j.copbio.2005.06.011>.
 63. Ishii T. 1991. Acetylation at O-2 of arabinofuranose residues in feruloylated arabinoxylan from bamboo shoot cell-walls. *Phytochemistry* 30: 2317–2320. [http://dx.doi.org/10.1016/0031-9422\(91\)83639-3](http://dx.doi.org/10.1016/0031-9422(91)83639-3).
 64. Ishii T, Tobita T. 1993. Structural characterization of feruloyl oligosaccharides from spinach-leaf cell walls. *Carbohydr. Res.* 248:179–190. [http://dx.doi.org/10.1016/0008-6215\(93\)84125-P](http://dx.doi.org/10.1016/0008-6215(93)84125-P).
 65. Laine C, Tamminen T, Hortling B. 2004. Carbohydrate structures in residual lignin-carbohydrate complexes of spruce and pine pulp. *Holzforschung* 58:611–621. <http://dx.doi.org/10.1515/HF.2004.115>.
 66. Sato S, Liu F, Koc H, Tien M. 2007. Expression analysis of extracellular proteins from *Phanerochaete chrysosporium* grown on different liquid and solid substrates. *Microbiology* 153:3023–3033. <http://dx.doi.org/10.1099/mic.0.2006/000513-0>.
 67. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. 2011. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* 7:461. <http://dx.doi.org/10.1038/msb.2010.116>.
 68. Henrissat B, Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 293:781–788.
 69. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6:41. <http://dx.doi.org/10.1186/1754-6834-6-41>.
 70. Beckham GT, Bomble YJ, Bayer EA, Himmel ME, Crowley MF. 2011. Applications of computational science for understanding enzymatic deconstruction of cellulose. *Curr. Opin. Biotechnol.* 22:231–238. <http://dx.doi.org/10.1016/j.copbio.2010.11.005>.
 71. Jovanovic I, Magnuson JK, Collart F, Robbertse B, Adney WS, Himmel ME, Baker SE. 2009. Fungal glycoside hydrolases for saccharification of lignocellulose: outlook for new discoveries fueled by genomics and functional studies. *Cellulose* 16:687–697. <http://dx.doi.org/10.1007/s10570-009-9307-z>.
 72. Harris PV, Welner D, McFarland KC, Re E, Navarro Poulsen JC, Brown K, Salbo R, Ding H, Vlasenko E, Merino S, Xu F, Cherry J, Larsen S, Lo Leggio L. 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49:3305–3316. <http://dx.doi.org/10.1021/bi100009p>.
 73. Langston JA, Shaghazi T, Abbate E, Xu F, Vlasenko E, Sweeney MD. 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.* 77:7007–7015. <http://dx.doi.org/10.1128/AEM.05815-11>.
 74. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JC, Johansen KS, Krogh KB, Jorgensen CI, Tovborg M, Anthonsen A, Tryfona T, Walter CP, Dupree P, Xu F, Davies GJ, Walton PH. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. U. S. A.* 108:15079–15084. <http://dx.doi.org/10.1073/pnas.1105776108>.
 75. Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Ejsink VG, Igarashi K, Samejima M, Stahlberg J, Horn SJ, Sandgren M. 2011. The putative endoglucanase PgCH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. *PLoS One* 6:e27807. <http://dx.doi.org/10.1371/journal.pone.0027807>.
 76. Sarkar P, Bosneaga E, Auer M. 2009. Plant cell walls throughout evolution: towards a molecular understanding of their design principles. *J. Exp. Bot.* 60:3615–3635. <http://dx.doi.org/10.1093/jxb/erp245>.
 77. Belien T, Van Campenhout S, Vanden Bosch A, Bourgois TM, Rombouts S, Robben J, Courtin CM, Delcour JA, Volckaert G. 2007. Engineering molecular recognition of endoxylanase enzymes and their inhibitors through phage display. *J. Mol. Recognit.* 20:103–112. <http://dx.doi.org/10.1002/jmr.818>.
 78. Wan CF, Chen WH, Chen CT, Chang MD, Lo LC, Li YK. 2007. Mutagenesis and mechanistic study of a glycoside hydrolase family 54 alpha-L-arabinofuranosidase from *Trichoderma koningii*. *Biochem. J.* 401:551–558. <http://dx.doi.org/10.1042/BJ20060717>.
 79. Vafiadis C, Topakas E, Biely P, Christakopoulos P. 2009. Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*. *FEMS Microbiol. Lett.* 296:178–184. <http://dx.doi.org/10.1111/j.1574-6968.2009.01631.x>.
 80. Ishida T, Yaoi K, Hirosaki A, Igarashi K, Samejima M. 2007. Substrate recognition by glycoside hydrolase family 74 xyloglucanase from the basidiomycete *Phanerochaete chrysosporium*. *FEBS J.* 274:5727–5736. <http://dx.doi.org/10.1111/j.1742-4658.2007.06092.x>.
 81. Carapito R, Imberty A, Jeltsch JM, Byrns SC, Tam PH, Lowary TL, Varrot A, Phalip V. 2009. Molecular basis of arabinobio-hydrolase activity in phytopathogenic fungi: crystal structure and catalytic mechanism of *Fusarium graminearum* GH93 exo-alpha-L-arabinanase. *J. Biol. Chem.* 284:12285–12296. <http://dx.doi.org/10.1074/jbc.M900439200>.
 82. Do BC, Dang TT, Berrin JG, Haltich D, To KA, Sigillot JC, Yamabhai M. 2009. Cloning, expression in *Pichia pastoris*, and characterization of a thermostable GH5 mannan endo-1,4-beta-mannosidase from *Aspergillus niger* BK01. *Microb. Cell Fact.* 8:59. <http://dx.doi.org/10.1186/1475-2859-8-59>.
 83. Taylor EJ, Gloster TM, Turkenburg JP, Vincent F, Brzozowski AM, Dupont C, Shareck F, Centeno MS, Prates JA, Puchart V, Ferreira LM, Fontes CM, Biely P, Davies GJ. 2006. Structure and activity of two metal ion-dependent acetylxyran esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. *J. Biol. Chem.* 281:10968–10975. <http://dx.doi.org/10.1074/jbc.M513066200>.
 84. Knoshaug EP, Selig MJ, Baker JO, Decker SR, Himmel ME, Adney WS. 2008. Heterologous expression of two ferulic acid esterases from *Penicillium funiculosum*. *Appl. Biochem. Biotechnol.* 146:79–87. <http://dx.doi.org/10.1007/s12010-007-8074-2>.
 85. Barker IJ, Petersen L, Reilly PJ. 2010. Mechanism of xylobiose hydrolysis by GH43 beta-xylosidase. *J. Phys. Chem. B* 114:15389–15393. <http://dx.doi.org/10.1021/jp107886e>.
 86. Prade RA. 1996. Xylanases: from biology to biotechnology. *Biotechnol. Genet. Eng. Rev.* 13:101–131. <http://dx.doi.org/10.1080/02648725.1996.10647925>.
 87. Schonknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Brautigam A, Baker BJ, Banfield JF, Garavito RM, Carr K, Wilkerson C, Rensing SA, Gagneul D, Dickenson NE, Oesterhelt C, Lercher MJ, Weber AP. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* 339:1207–1210. <http://dx.doi.org/10.1126/science.1231707>.
 88. Beckham GT, Matthews JF, Bomble YJ, Bu L, Adney WS, Himmel ME, Nimlos MR, Crowley MF. 2010. Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. *J. Phys. Chem. B* 114:1447–1453. <http://dx.doi.org/10.1021/jp908810a>.
 89. Bu L, Beckham GT, Shirts MR, Nimlos MR, Adney WS, Himmel ME, Crowley MF. 2011. Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. *J. Biol. Chem.* 286:18161–18169. <http://dx.doi.org/10.1074/jbc.M110.212076>.

90. Igarashi K, Uchihashi T, Koivula A, Wada M, Kimura S, Okamoto T, Penttila M, Ando T, Samejima M. 2011. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* 333:1279–1282. <http://dx.doi.org/10.1126/science.1208386>.
91. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, Goedegebur F, Houfek TD, England GJ, Kelley AS, Meerman HJ, Mitchell T, Mitchinson C, Olivares HA, Teunissen PJ, Yao J, Ward M. 2003. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J. Biol. Chem.* 278:31988–31997. <http://dx.doi.org/10.1074/jbc.M304750200>.
92. Okada H, Tada K, Sekiya T, Yokoyama K, Takahashi A, Tohda H, Kumagai H, Morikawa Y. 1998. Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endogluconanase from *Trichoderma reesei* QM9414. *Appl. Environ. Microbiol.* 64:555–563.
93. Saloheimo M, Lehtovaara P, Penttila M, Teeri TT, Stahlberg J, Johansson G, Pettersson G, Claeysens M, Tomme P, Knowles JK. 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene* 63:11–22. [http://dx.doi.org/10.1016/0378-1119\(88\)90541-0](http://dx.doi.org/10.1016/0378-1119(88)90541-0).
94. Saloheimo M, Nakari-Setala T, Tenkanen M, Penttila M. 1997. cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *Eur. J. Biochem.* 249:584–591. <http://dx.doi.org/10.1111/j.1432-1033.1997.00584.x>.
95. Teeri TT, Lehtovaara P, Kauppinen S, Salovuori I, Knowles J. 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. *Gene* 51:43–52. [http://dx.doi.org/10.1016/0378-1119\(87\)90472-0](http://dx.doi.org/10.1016/0378-1119(87)90472-0).
96. Barnett CC, Berka RM, Fowler T. 1991. Cloning and amplification of the gene encoding an extracellular beta-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *BioTechnology* 9:562–567. <http://dx.doi.org/10.1038/nbt0691-562>.
97. Takashima S, Nakamura A, Hidaka M, Masaki H, Uozumi T. 1999. Molecular cloning and expression of the novel fungal beta-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. *J. Biochem.* 125: 728–736. <http://dx.doi.org/10.1093/oxfordjournals.jbchem.a022343>.
98. Wang H, Squina F, Segato F, Mort A, Lee D, Pappan K, Prade R. 2011. High-temperature enzymatic breakdown of cellulose. *Appl. Environ. Microbiol.* 77:5199–5206. <http://dx.doi.org/10.1128/AEM.00199-11>.
99. Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* 3:853–859. [http://dx.doi.org/10.1016/S0969-2126\(01\)00220-9](http://dx.doi.org/10.1016/S0969-2126(01)00220-9).
100. Segato F, Damasio AR, Goncalves TA, Murakami MT, Squina FM, Polizeli M, Mort AJ, Prade RA. 2012. Two structurally discrete GH7 cellobiohydrolases compete for the same cellulosic substrate fiber. *BioTechnol. Biofuels* 5:21. <http://dx.doi.org/10.1186/1754-6834-5-21>.
101. Varnai A, Siika-Aho M, Viikari L. 2013. Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnol. Biofuels* 6:30. <http://dx.doi.org/10.1186/1754-6834-6-30>.
102. Berghem LE, Pettersson LG. 1973. The mechanism of enzymatic cellulose degradation. Purification of a cellulolytic enzyme from *Trichoderma viride* active on highly ordered cellulose. *Eur. J. Biochem.* 37:21–30.
103. Eriksson K-E, Pettersson B. 1975. Extracellular enzyme system utilized by the fungus *Sporotrichum pulverulentum* (*Chrysosporium lignorum*) for the breakdown of cellulose. 3. Purification and physico-chemical characterization of an exo-1,4-beta-glucanase. *Eur. J. Biochem.* 51:213–218.
104. Takahashi M, Takahashi H, Nakano Y, Konishi T, Terauchi R, Takeda T. 2010. Characterization of a cellobiohydrolase (MoCel6A) produced by *Magnaporthe oryzae*. *Appl. Environ. Microbiol.* 76:6583–6590. <http://dx.doi.org/10.1128/AEM.00618-10>.
105. Knowles JKC, Lentovaara P, Murray M, Sinnott ML. 1988. Stereochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. *J. Chem. Soc. Chem. Commun.* 1988:1401–1402.
106. Rouvinen J, Bergfors T, Teeri T, Knowles J, Jones T. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 249:380–386. <http://dx.doi.org/10.1126/science.2377893>.
107. Varrot A, Frandsen TP, Driguez H, Davies GJ. 2002. Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio-[beta]-cellobioside, at 1.9 Å. *Acta Crystallogr. D Biol. Crystallogr.* 58: 2201–2204. <http://dx.doi.org/10.1107/S0907444902017006>.
108. Zou J-Y, Kleywegt GJ, Stahlberg J, Driguez H, Nerinckx W, Claeysens M, Koivula A, Teeri TT, Jones TA. 1999. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure* 7:1035–1045. [http://dx.doi.org/10.1016/S0969-2126\(99\)80171-3](http://dx.doi.org/10.1016/S0969-2126(99)80171-3).
109. Divne C, Stahlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles J, Teeri T, Jones T. 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265:524–528. <http://dx.doi.org/10.1126/science.8036495>.
110. Stahlberg J, Divne C, Koivula A, Piens K, Claeysens M, Teeri TT, Jones TA. 1996. Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* 264:337–349. <http://dx.doi.org/10.1006/jmbi.1996.0644>.
111. Igarashi K, Koivula A, Wada M, Kimura S, Penttila M, Samejima M. 2009. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *J. Biol. Chem.* 284:36186–36190. <http://dx.doi.org/10.1074/jbc.M109.034611>.
112. Kurasin M, Välijamäe P. 2011. Processivity of cellobiohydrolases is limited by the substrate. *J. Biol. Chem.* 286:169–177. <http://dx.doi.org/10.1074/jbc.M110.161059>.
113. Liu D, Zhang R, Yang X, Xu Y, Tang Z, Tian W, Shen Q. 2011. Expression, purification and characterization of two thermostable endoglucanases cloned from a lignocellulosic decomposing fungi *Aspergillus fumigatus* Z5 isolated from compost. *Protein Expr. Purif.* 79:176–186. <http://dx.doi.org/10.1016/j.pep.2011.06.008>.
114. Hara Y, Shimoi HYH, Ito K. 2003. Cloning and sequence analyses of endoglucanase genes from an industrial fungus, *Aspergillus kawachii*. *Biosci. Biotechnol. Biochem.* 67:2010–2013. <http://dx.doi.org/10.1271/bbb.67.2010>.
115. Lee TM, Farrow MF, Arnold FH, Mayo SL. 2011. A structural study of *Hypocrea jecorina* Cel5A. *Protein Sci.* 20:1935–1940. <http://dx.doi.org/10.1002/pro.730>.
116. Lo Leggio L, Larsen S. 2002. The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett.* 523:103–108. [http://dx.doi.org/10.1016/S0014-5793\(02\)02954-X](http://dx.doi.org/10.1016/S0014-5793(02)02954-X).
117. Mertz B, Gu X, Reilly PJ. 2009. Analysis of functional divergence within two structurally related glycoside hydrolase families. *Biopolymers* 91: 478–495. <http://dx.doi.org/10.1002/bip.21154>.
118. Davies GJ, Ducros V, Lewis RJ, Borchert TV, Schulein M. 1997. Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites. *J. Biotechnol.* 57:91–100. [http://dx.doi.org/10.1016/S0168-1656\(97\)00092-8](http://dx.doi.org/10.1016/S0168-1656(97)00092-8).
119. Kleywegt GJ, Zou JY, Divne C, Davies GJ, Sinning I, Stahlberg J, Reinikainen T, Srisodsuk M, Teeri TT, Jones TA. 1997. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *J. Mol. Biol.* 272:383–397. <http://dx.doi.org/10.1006/jmbi.1997.1243>.
120. Sulzenbacher G, Driguez H, Henrissat B, Schulein M, Davies GJ. 1996. Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* 35:15280–15287. <http://dx.doi.org/10.1021/bi961946h>.
121. Sandgren M, Gualfetti PJ, Shaw A, Gross LS, Saldajeno M, Day AG, Jones TA, Mitchinson C. 2003. Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. *Protein Sci.* 12:848–860. <http://dx.doi.org/10.1110/ps.0237703>.
122. Sandgren M, Shaw A, Ropp TH, Wu S, Bott R, Cameron AD, Stahlberg J, Mitchinson C, Jones TA. 2001. The X-ray crystal structure of the *Trichoderma reesei* family 12 endoglucanase 3, Cel12A, at 1.9 Å resolution. *J. Mol. Biol.* 308:295–310. <http://dx.doi.org/10.1006/jmbi.2001.4583>.
123. Khademi S, Zhang D, Swanson SM, Wartenberg A, Witte K, Meyer EF. 2002. Determination of the structure of an endoglucanase from *Aspergillus niger* and its mode of inhibition by palladium chloride. *Acta Crystallogr. D Biol. Crystallogr.* 58:660–667. <http://dx.doi.org/10.1107/S0907444902003360>.
124. Vlasenko E, Schulein M, Cherry J, Xu F. 2010. Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresour. Technol.* 101: 2405–2411. <http://dx.doi.org/10.1016/j.biortech.2009.11.057>.
125. Aspeborg H, Coutinho PM, Wang Y, Brumer H, 3rd, Henrissat B. 2012. Evolution, substrate specificity and subfamily classification of gly-

- coside hydrolase family 5 (GH5). *BMC Evol. Biol.* 12:186. <http://dx.doi.org/10.1186/1471-2148-12-186>.
126. Wu M, Beckham GT, Larsson AM, Ishida T, Kim S, Payne CM, Himmel ME, Crowley MF, Horn SJ, Westereng B, Igarashi K, Samejima M, Stahlberg J, Eijsink VG, Sandgren M. 2013. Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the Basidiomycota fungus *Phanerochaete chrysosporium*. *J. Biol. Chem.* 288:12828–12839. <http://dx.doi.org/10.1074/jbc.M113.459396>.
 127. Beeson WT, Phillips CM, Cate JH, Marletta MA. 2012. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J. Am. Chem. Soc.* 134:890–892. <http://dx.doi.org/10.1021/ja210657t>.
 128. Dimarogona M, Topakas E, Christakopoulos P. 2013. Recalcitrant polysaccharide degradation by novel oxidative biocatalysts. *Appl. Microbiol. Biotechnol.* 97:8455–8465. <http://dx.doi.org/10.1007/s00253-013-5197-y>.
 129. Li X, Beeson WT, 4th, Phillips CM, Marletta MA, Cate JH. 2012. Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* 20:1051–1061. <http://dx.doi.org/10.1016/j.str.2012.04.002>.
 130. Karkehabadi S, Hansson H, Kim S, Piens K, Mitchinson C, Sandgren M. 2008. The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocreaf jecorina*, at 1.6 Å resolution. *J. Mol. Biol.* 383: 144–154. <http://dx.doi.org/10.1016/j.jmb.2008.08.016>.
 131. Igarashi K, Yoshida M, Matsumura H, Nakamura N, Ohno H, Samejima M, Nishino T. 2005. Electron transfer chain reaction of the extracellular flavocytochrome cellobiose dehydrogenase from the basidiomycete *Phanerochaete chrysosporium*. *FEBS J.* 272:2869–2877. <http://dx.doi.org/10.1111/j.1742-4658.2005.04707.x>.
 132. Mansfield SD, De Jong E, Saddler JN. 1997. Cellobiose dehydrogenase, an active agent in cellulose depolymerization. *Appl. Environ. Microbiol.* 63:3804–3809.
 133. Zamocky M, Hallberg M, Ludwig R, Divne C, Haltrich D. 2004. Ancestral gene fusion in cellobiose dehydrogenases reflects a specific evolution of GMC oxidoreductases in fungi. *Gene* 338:1–14. <http://dx.doi.org/10.1016/j.gene.2004.04.025>.
 134. Westermark U, Eriksson KE. 1975. Purification and properties of cellobiose: quinone oxidoreductase from *Sporotrichum pulverulentum*. *Acta Chem. Scand. B* 29:419–424.
 135. Ekwe E, Morgenstern I, Tsang A, Storms R, Powlowski J. 2013. Non-hydrolytic cellulose active proteins: research progress and potential application in biorefineries. *Ind. Biotechnol.* 9:123–131. <http://dx.doi.org/10.1089/ind.2013.0010>.
 136. Beeson WT, Iavarone AT, Hausmann CD, Cate JH, Marletta MA. 2011. Extracellular aldonolactonase from *Myceliophthora thermophila*. *Appl. Environ. Microbiol.* 77:650–656. <http://dx.doi.org/10.1128/AEM.01922-10>.
 137. Canlevscini G, Borer P, Dreyer JL. 1991. Cellobiose dehydrogenases of *Sporotrichum* (*Chrysosporium*) thermophile. *Eur. J. Biochem.* 198:43–52. <http://dx.doi.org/10.1111/j.1432-1033.1991.tb15984.x>.
 138. Mason MG, Nicholls P, Wilson MT. 2003. Rotting by radicals—the role of cellobiose oxidoreductase? *Biochem. Soc. Trans.* 31:1335–1336. <http://dx.doi.org/10.1042/BST0311335>.
 139. Zamocky M, Ludwig R, Peterbauer C, Hallberg BM, Divne C, Nicholls P, Haltrich D. 2006. Cellobiose dehydrogenase—a flavocytochrome from wood-degrading, phytopathogenic and saprotropic fungi. *Curr. Protein Pept. Sci.* 7:255–280. <http://dx.doi.org/10.2174/138920306777452367>.
 140. Hemsworth GR, Davies GJ, Walton PH. 2013. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Curr. Opin. Struct. Biol.* 23:660–668. <http://dx.doi.org/10.1016/j.sbi.2013.05.006>.
 141. Zamocky M, Schumann C, Sygmund C, O'Callaghan J, Dobson AD, Ludwig R, Haltrich D, Peterbauer CK. 2008. Cloning, sequence analysis and heterologous expression in *Pichia pastoris* of a gene encoding a thermostable cellobiose dehydrogenase from *Myriococcum thermophilum*. *Protein Expr. Purif.* 59:258–265. <http://dx.doi.org/10.1016/j.pep.2008.02.007>.
 142. Ludwig R, Harreither W, Tasca F, Gorton L. 2010. Cellobiose dehydrogenase: a versatile catalyst for electrochemical applications. *Chembioschem* 11:2674–2697. <http://dx.doi.org/10.1002/cphc.201000216>.
 143. Subramanian SS, Nagalla SR, Renganathan V. 1999. Cloning and characterization of a thermostable cellobiose dehydrogenase from *Sporotrichum thermophile*. *Arch. Biochem. Biophys.* 365:223–230. <http://dx.doi.org/10.1006/abbi.1999.1152>.
 144. Hallberg BM, Henriksson G, Pettersson G, Divne C. 2002. Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *J. Mol. Biol.* 315:421–434. <http://dx.doi.org/10.1006/jmbi.2001.5246>.
 145. Bhatia Y, Mishra S, Bisaria VS. 2002. Microbial beta-glucosidases: cloning, properties, and applications. *Crit. Rev. Biotechnol.* 22:375–407. <http://dx.doi.org/10.1080/07388550290789568>.
 146. Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280:309–316.
 147. Henrissat B, Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7:637–644. [http://dx.doi.org/10.1016/S0959-4420\(97\)80072-3](http://dx.doi.org/10.1016/S0959-4420(97)80072-3).
 148. Jeng WY, Wang NC, Lin MH, Lin CT, Liaw YC, Chang WJ, Liu CI, Liang PH, Wang AH. 2011. Structural and functional analysis of three beta-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *J. Struct. Biol.* 173:46–56. <http://dx.doi.org/10.1016/j.jsb.2010.07.008>.
 149. Suzuki K, Sumitani J, Nam YW, Nishimaki T, Tani S, Wakagi T, Kawaguchi T, Fushinobu S. 2013. Crystal structures of glycoside hydrolase family 3 beta-glucosidase 1 from *Aspergillus aculeatus*. *Biochem. J.* 452:211–221. <http://dx.doi.org/10.1042/BJ20130054>.
 150. Gebler J, Gilkes NR, Claeysens M, Wilson DB, Beguin P, Wakarchuk WW, Kilburn DG, Miller RC, Jr, Warren RA, Withers SG. 1992. Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases. *J. Biol. Chem.* 267:12559–12561.
 151. Goncalves TA, Damasio AR, Segato F, Alvarez TM, Bragatto J, Brenelli LB, Citadini AP, Murakami MT, Ruller R, Paes Leme AF, Prade RA, Squina FM. 2012. Functional characterization and synergic action of fungal xylanase and arabinofuranosidase for production of xylooligosaccharides. *Bioresour. Technol.* 119:293–299. <http://dx.doi.org/10.1016/j.biotech.2012.05.062>.
 152. Zechel DL, Konermann L, Withers SG, Douglas DJ. 1998. Pre-steady state kinetic analysis of an enzymatic reaction monitored by time-resolved electrospray ionization mass spectrometry. *Biochemistry* 37: 7664–7669. <http://dx.doi.org/10.1021/bi980445o>.
 153. MacLeod AM, Lindhorst T, Withers SG, Warren RA. 1994. The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry* 33:6371–6376. <http://dx.doi.org/10.1021/bi00186a042>.
 154. Damásio ARDL, Silva TM, dos Reis Almeida FB, Squina FM, Ribeiro DA, Leme AFP, Segato F, Prade RA, Jorge JA, Terenzi HF, de Moraes Polizeli MDLT. 2011. Heterologous expression of an *Aspergillus niveus* xylanase GH11 in *Aspergillus nidulans* and its characterization and application. *Process Biochem.* 46:1236–1242. <http://dx.doi.org/10.1016/j.procbio.2011.01.027>.
 155. Squina FM, Mort AJ, Decker SR, Prade RA. 2009. Xylan decomposition by *Aspergillus clavatus* endo-xylanase. *Protein Expr. Purif.* 68:65–71. <http://dx.doi.org/10.1016/j.pep.2009.06.014>.
 156. Teplitsky A, Mechaly A, Stojanoff V, Sainz G, Golan G, Feinberg H, Gilboa R, Reiland V, Zolotnitsky G, Shalom D, Thompson A, Shoham Y, Shoham G. 2004. Structure determination of the extracellular xylanase from *Geobacillus stearothermophilus* by selenomethionyl MAD phasing. *Acta Crystallogr. D Biol. Crystallogr.* 60:836–848. <http://dx.doi.org/10.1107/S090744904004123>.
 157. Andre-Leroux G, Berrin JG, Georis J, Arnaut F, Juge N. 2008. Structure-based mutagenesis of *Penicillium griseofulvum* xylanase using computational design. *Proteins* 72:1298–1307. <http://dx.doi.org/10.1002/prot.22029>.
 158. Gallardo O, Fernandez-Fernandez M, Valls C, Valenzuela SV, Roncero MB, Vidal T, Diaz P, Pastor FI. 2010. Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. *Appl. Environ. Microbiol.* 76:6290–6294. <http://dx.doi.org/10.1128/AEM.00871-10>.
 159. Tenkanen M, Vrsanska M, Siika-aho M, Wong DW, Puchart V, Penttila M, Saloheimo M, Biely P. 2013. Xylanase XYN IV from *Trichoderma reesei* showing exo- and endo-xylanase activity. *FEBS J.* 280:285–301. <http://dx.doi.org/10.1111/febs.12069>.
 160. Rasmussen LE, Sorensen HR, Vind J, Vikso-Nielsen A. 2006. Mode of action and properties of the beta-xylosidases from *Talaromyces emersonii* and *Trichoderma reesei*. *Biotechnol. Bioeng.* 94:869–876. <http://dx.doi.org/10.1002/bit.20908>.

161. Eneyeskaya EV, Ivanen DR, Bobrov KS, Isaeva-Ivanova LS, Shabalina KA, Savel'ev AN, Golubev AM, Kulminskaya AA. 2007. Biochemical and kinetic analysis of the GH3 family beta-xylosidase from *Aspergillus awamori* X-100. *Arch. Biochem. Biophys.* 457:225–234. <http://dx.doi.org/10.1016/j.abb.2006.10.024>.
162. Gomez M, Isorna P, Rojo M, Estrada P. 2001. Chemical mechanism of beta-xylosidase from *Trichoderma reesei* QM 9414: pH-dependence of kinetic parameters. *Biochimie* 83:961–967. [http://dx.doi.org/10.1016/S0300-9084\(01\)01341-4](http://dx.doi.org/10.1016/S0300-9084(01)01341-4).
163. Kiss T, Erdei A, Kiss L. 2002. Investigation of the active site of the extracellular beta-D-xylosidase from *Aspergillus carbonarius*. *Arch. Biochem. Biophys.* 399:188–194. <http://dx.doi.org/10.1006/abbi.2002.2753>.
164. Brux C, Ben-David A, Shallom-Shezifi D, Leon M, Niefeld K, Shoham G, Shoham Y, Schomburg D. 2006. The structure of an inverting GH43 beta-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J. Mol. Biol.* 359:97–109. <http://dx.doi.org/10.1016/j.jmb.2006.03.005>.
165. Vandermarliere E, Bourgois TM, Winn MD, van Campenhout S, Volckaert G, Delcour JA, Strelkov SV, Rabijns A, Courtin CM. 2009. Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem. J.* 418:39–47. <http://dx.doi.org/10.1042/BJ20081256>.
166. Songsiriritthigul C, Buranabanyat B, Haltrich D, Yamabhai M. 2010. Efficient recombinant expression and secretion of a thermostable GH26 mannan endo-1,4-beta-mannosidase from *Bacillus licheniformis* in *Escherichia coli*. *Microb. Cell Fact.* 9:20. <http://dx.doi.org/10.1186/1475-2859-9-20>.
167. Bolam DN, Hughes N, Virden R, Lakey JH, Hazlewood GP, Henrissat B, Braithwaite KL, Gilbert HJ. 1996. Mannanase A from *Pseudomonas fluorescens* ssp. *cellulosa* is a retaining glycosyl hydrolase in which E212 and E320 are the putative catalytic residues. *Biochemistry* 35:16195–16204. <http://dx.doi.org/10.1021/bi961866d>.
168. Ducros VM, Zechel DL, Murshudov GN, Gilbert HJ, Szabo L, Stoll D, Withers SG, Davies GJ. 2002. Substrate distortion by a beta-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a B(2,5) conformation for the transition state. *Angew. Chem. Int. Ed. Engl.* 41:2824–2827. [http://dx.doi.org/10.1002/1521-3773\(20020802\)41:15<2824::AID-ANIE2824>3.0.CO;2-G](http://dx.doi.org/10.1002/1521-3773(20020802)41:15<2824::AID-ANIE2824>3.0.CO;2-G).
169. Hilge M, Gloor SM, Rypniewski W, Sauer O, Heightman TD, Zimmermann W, Winterhalter K, Piontek K. 1998. High-resolution native and complex structures of thermostable beta-mannanase from *Thermomonospora fusca*—substrate specificity in glycosyl hydrolase family 5. *Structure* 6:1433–1444. [http://dx.doi.org/10.1016/S0969-2126\(98\)00142-7](http://dx.doi.org/10.1016/S0969-2126(98)00142-7).
170. Sabini E, Schubert H, Murshudov G, Wilson KS, Siika-Aho M, Penttila M. 2000. The three-dimensional structure of a *Trichoderma reesei* beta-mannanase from glycoside hydrolase family 5. *Acta Crystallogr. D Biol. Crystallogr.* 56:3–13. <http://dx.doi.org/10.1107/S0907444999013943>.
171. Seibold B, Metz B. 2011. Fungal arabinan and L-arabinose metabolism. *Appl. Microbiol. Biotechnol.* 89:1665–1673. <http://dx.doi.org/10.1007/s00253-010-3071-8>.
172. Puls J. 1997. Chemistry and biochemistry of hemicelluloses: relationship between hemicellulose structure and enzymes required for hydrolysis. *Macromol. Symp.* 120:183–196. <http://dx.doi.org/10.1002/masy.19971200119>.
173. Beldman G, Searle-van Leeuwen MJF, De Ruiter GA, Siliha HA, Voragen AGJ. 1993. Degradation of arabinans by arabinanases from *Aspergillus aculeatus* and *Aspergillus niger*. *Carbohydr. Polymers* 20:159–168. [http://dx.doi.org/10.1016/0144-8617\(93\)90146-U](http://dx.doi.org/10.1016/0144-8617(93)90146-U).
174. Kaneko S, Arimoto M, Ohba M, Kobayashi H, Ishii T, Kusakabe I. 1998. Purification and substrate specificities of two alpha-L-arabinofuranosidases from *Aspergillus awamori* IFO 4033. *Appl. Environ. Microbiol.* 64:4021–4027.
175. Saha BC. 2000. Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. *Biotechnol. Adv.* 18:403–423. [http://dx.doi.org/10.1016/S0734-9750\(00\)00044-6](http://dx.doi.org/10.1016/S0734-9750(00)00044-6).
176. Miyanaga A, Koseki T, Matsuzawa H, Wakagi T, Shoun H, Fushinobu S. 2004. Crystal structure of a family 54 alpha-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. *J. Biol. Chem.* 279:44907–44914. <http://dx.doi.org/10.1074/jbc.M405390200>.
177. dos Santos CR, Squina FM, Navarro AM, Oldiges DP, Leme AF, Ruller R, Mort AJ, Prade R, Murakami MT. 2011. Functional and biophysical characterization of a hyperthermstable GH51 alpha-L-arabinofuranosidase from *Thermotoga petrophila*. *Biotechnol. Lett.* 33:131–137. <http://dx.doi.org/10.1007/s10529-010-0409-3>.
178. Taylor EJ, Smith NL, Turkenburg JP, D'Souza S, Gilbert HJ, Davies GJ. 2006. Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum*. *Biochem. J.* 395:31–37. <http://dx.doi.org/10.1042/BJ20051780>.
179. Damásio ARDL, Pessela BC, Mateo C, Segato F, Prade RA, Guisan JM, de Moraes Polizeli MDLT. 2012. Immobilization of a recombinant endo-1,5-arabinanase secreted by *Aspergillus nidulans* strain A773. *J. Mol. Catal. B Enzym.* 77:39–45. <http://dx.doi.org/10.1016/j.molcatb.2012.01.002>.
180. Naumoff DG. 2001. Beta-fructosidase superfamily: homology with some alpha-L-arabinases and beta-D-xylosidases. *Proteins* 42:66–76. [http://dx.doi.org/10.1002/1097-0134\(20010101\)42:1<66::AID-PROT70>3.0.CO;2-4](http://dx.doi.org/10.1002/1097-0134(20010101)42:1<66::AID-PROT70>3.0.CO;2-4).
181. Proctor MR, Taylor EJ, Nurizzo D, Turkenburg JP, Lloyd RM, Vardakou M, Davies GJ, Gilbert HJ. 2005. Tailored catalysts for plant cell-wall degradation: redesigning the exo/endo preference of *Cellvibrio japonicus* arabinanase 43A. *Proc. Natl. Acad. Sci. U. S. A.* 102:2697–2702. <http://dx.doi.org/10.1073/pnas.0500051102>.
182. de Sanctis D, Inacio JM, Lindley PF, de Sa-Nogueira I, Bento I. 2010. New evidence for the role of calcium in the glycosidase reaction of GH43 arabinanases. *FEBS J.* 277:4562–4574. <http://dx.doi.org/10.1111/j.1742-4658.2010.07870.x>.
183. Nurizzo D, Turkenburg JP, Charnock SJ, Roberts SM, Dodson EJ, McKie VA, Taylor EJ, Gilbert HJ, Davies GJ. 2002. *Cellvibrio japonicus* alpha-L-arabinanase 43A has a novel five-blade beta-propeller fold. *Nat. Struct. Biol.* 9:665–668. <http://dx.doi.org/10.1038/nsb835>.
184. Gogabe Y, Kitatani T, Yamaguchi A, Kinoshita T, Adachi H, Takano K, Inoue T, Mori Y, Matsumura H, Sakamoto T, Tada T. 2011. High-resolution structure of exo-arabinanase from *Penicillium chrysogenum*. *Acta Crystallogr. D Biol. Crystallogr.* 67:415–422. <http://dx.doi.org/10.1107/S0907444911006299>.
185. Ovodov YS. 1975. Structural chemistry of plant glyuronoglycans. *Pure Appl. Chem.* 42:351–369.
186. Delattre C, Michaud P, Keller C, Elboutachfaïti R, Beven L, Courtois B, Courtois J. 2006. Purification and characterization of a novel glucuronan lyase from *Trichoderma* sp. GL2. *Appl. Microbiol. Biotechnol.* 70:437–443. <http://dx.doi.org/10.1007/s00253-005-0077-8>.
187. Dick-Perez M, Zhang Y, Hayes J, Salazar A, Zabotina OA, Hong M. 2011. Structure and interactions of plant cell-wall polysaccharides by two- and three-dimensional magic-angle-spinning solid-state NMR. *Biochemistry* 50:989–1000. <http://dx.doi.org/10.1021/bi101795q>.
188. Gloster TM, Ibatullin FM, Macauley K, Eklof JM, Roberts S, Turkenburg JP, Bjornvad ME, Jorgensen PL, Danielsen S, Johansen KS, Borchart TV, Wilson KS, Brumer H, Davies GJ. 2007. Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J. Biol. Chem.* 282:19177–19189. <http://dx.doi.org/10.1074/jbc.M700224200>.
189. Damásio AR, Rubio MV, Oliveira LC, Segato F, Dias BA, Citadini AP, Paixão DA, Squina FM. 2014. Understanding the function of conserved variations in the catalytic loops of fungal glycoside hydrolase family 12. *Biotechnol. Bioeng.* 111:1494–1505. <http://dx.doi.org/10.1002/bit.25209>.
190. Grishutin SG, Gusakov AV, Markov AV, Ustinov BB, Semenova MV, Sinitsyn AP. 2004. Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. *Biochim. Biophys. Acta* 1674:268–281. <http://dx.doi.org/10.1016/j.bbagen.2004.07.001>.
191. Powłowski J, Mahajan S, Schapira M, Master ER. 2009. Substrate recognition and hydrolysis by a fungal xyloglucan-specific family 12 hydrolase. *Carbohydr. Res.* 344:1175–1179. <http://dx.doi.org/10.1016/j.carres.2009.04.020>.
192. Damásio AR, Ribeiro LF, Ribeiro LF, Furtado GP, Segato F, Almeida FB, Crivellari AC, Buckeridge MS, Souza TA, Murakami MT, Ward RJ, Prade RA, Polizeli ML. 2012. Functional characterization and oligomerization of a recombinant xyloglucan-specific endo-beta-1,4-glucanase (GH12) from *Aspergillus niveus*. *Biochim. Biophys. Acta* 1824:461–467. <http://dx.doi.org/10.1016/j.bbapap.2011.12.005>.
193. Pauly M, Albersheim P, Darvill A, York WS. 1999. Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *Plant J.* 20:629–639. <http://dx.doi.org/10.1046/j.1365-313X.1999.00630.x>.
194. Pauly M, Andersen LN, Kauppinen S, Kofod LV, York WS, Albersheim P, Darvill A. 1999. A xyloglucan-specific endo-beta-1,4-glucanase from *Aspergillus aculeatus*: expression cloning in yeast, puri-

- fication and characterization of the recombinant enzyme. *Glycobiology* 9:93–100. <http://dx.doi.org/10.1093/glycob/9.1.93>.
195. Bauer S, Vasu P, Mort AJ, Somerville CR. 2005. Cloning, expression, and characterization of an oligoxyloglucan reducing end-specific xyloglucanohydrodrolase from *Aspergillus nidulans*. *Carbohydr. Res.* 340: 2590–2597. <http://dx.doi.org/10.1016/j.carres.2005.09.014>.
 196. Yaoi K, Mitsuishi Y. 2002. Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. *J. Biol. Chem.* 277:48276–48281. <http://dx.doi.org/10.1074/jbc.M208443200>.
 197. Fry SC, York WS, Albersheim P, Darvill A, Hayashi T, Joseleau J-P, Kato Y, Llorenç EP, MacLachlan GA, McNeil M, Mort AJ, Grant Reid JS, Seitz HU, Selvendran RR, Voragen AGJ, White AR. 1993. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Phyiol. Plant.* 89:1–3.
 198. Yaoi K, Kondo H, Hirosaki A, Noro N, Sugimoto H, Tsuda S, Mitsuishi Y, Miyazaki K. 2007. The structural basis for the exo-mode of action in GH74 oligoxyloglucan reducing end-specific cellobiohydrolase. *J. Mol. Biol.* 370:53–62. <http://dx.doi.org/10.1016/j.jmb.2007.04.035>.
 199. Yaoi K, Kondo H, Noro N, Suzuki M, Tsuda S, Mitsuishi Y. 2004. Tandem repeat of a seven-bladed beta-propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. *Structure* 12:1209–1217. <http://dx.doi.org/10.1016/j.str.2004.04.020>.
 200. Yaoi K, Kondo H, Suzuki M, Noro N, Tsuda S, Mitsuishi Y. 2003. Crystallization and preliminary X-ray crystallographic study on a xyloglucan-specific exo-beta-glycosidase, oligoxyloglucan reducing-end specific cellobiohydrolase. *Acta Crystallogr. D Biol. Crystallogr.* 59:1838–1839. <http://dx.doi.org/10.1107/S0907444903017128>.
 201. Matsuzawa T, Saito Y, Yaoi K. 2014. Key amino acid residues for the endo-processive activity of GH74 xyloglucanase. *FEBS Lett.* 588:1731–1738. <http://dx.doi.org/10.1016/j.febslet.2014.03.023>.
 202. Murphy C, Powłowski J, Wu M, Butler G, Tsang A. 2011. Curation of characterized glycoside hydrolases of fungal origin. *Database (Oxford)* 2011:bar020. <http://dx.doi.org/10.1093/database/bar020>.
 203. Liu D, Li J, Zhao S, Zhang R, Wang M, Miao Y, Shen Y, Shen Q. 2013. Secretome diversity and quantitative analysis of cellulolytic *Aspergillus fumigatus* Z5 in the presence of different carbon sources. *Biotechnol. Biofuels* 6:149. <http://dx.doi.org/10.1186/1754-6834-6-149>.
 204. Ribeiro DA, Cota J, Alvarez TM, Bruchli F, Bragato J, Pereira BM, Pauletti BA, Jackson G, Pimenta MT, Murakami MT, Camassola M, Ruller R, Dillon AJ, Pradella JG, Paes Leme AF, Squina FM. 2012. The *Penicillium echinulatum* secretome on sugar cane bagasse. *PLoS One* 7:e50571. <http://dx.doi.org/10.1371/journal.pone.0050571>.
 205. Koseki T, Miwa Y, Akao T, Akita O, Hashizume K. 2006. An *Aspergillus oryzae* acetyl xylan esterase: molecular cloning and characteristics of recombinant enzyme expressed in *Pichia pastoris*. *J. Biotechnol.* 121: 381–389. <http://dx.doi.org/10.1016/j.biote.2005.07.015>.
 206. Koseki T, Miwa Y, Fushinobu S, Hashizume K. 2005. Biochemical characterization of recombinant acetyl xylan esterase from *Aspergillus awamori* expressed in *Pichia pastoris*: mutational analysis of catalytic residues. *Biochim. Biophys. Acta* 1749:7–13. <http://dx.doi.org/10.1016/j.bbapap.2005.01.009>.
 207. Margolles-Clark E, Tenkanen M, Soderlund H, Penttila M. 1996. Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain. *Eur. J. Biochem.* 237:553–560. <http://dx.doi.org/10.1111/j.1432-1033.1996.0553p.x>.
 208. Li XL, Skory CD, Cotta MA, Puchart V, Biely P. 2008. Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene. *Appl. Environ. Microbiol.* 74:7482–7489. <http://dx.doi.org/10.1128/AEM.00807-08>.
 209. Poutanen K, Sundberg M. 1988. An acetyl esterase of *Trichoderma reesei* and its role in the hydrolysis of acetyl xylans. *Appl. Microbiol. Biotechnol.* 28:419–424. <http://dx.doi.org/10.1007/BF00268207>.
 210. Poutanen K, Sundberg M, Korte H, Puls J. 1990. Deacetylation of xylans by acetyl esterases of *Trichoderma reesei*. *Appl. Microbiol. Biotechnol.* 33:506–510.
 211. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37: D233–D238. <http://dx.doi.org/10.1093/nar/gkn663>.
 212. Altaner C. 2003. Regioselective deacetylation of cellulose acetates by acetyl xylan esterases of different CE-families. *J. Biotechnol.* 105:95–104. [http://dx.doi.org/10.1016/S0168-1656\(03\)00187-1](http://dx.doi.org/10.1016/S0168-1656(03)00187-1).
 213. Tenkanen M, Eyzaguirre J, Isoniemi R, Faulds CB, Biely P. 2003. Comparison of catalytic properties of acetyl xylan esterases from three carbohydrate esterase families. *ACS Symp. Ser.* 855:211–229. <http://dx.doi.org/10.1021/bk-2003-0855.ch013>.
 214. Hakulinen N, Tenkanen M, Rouvinen J. 2000. Three-dimensional structure of the catalytic core of acetyl xylan esterase from *Trichoderma reesei*: insights into the deacetylation mechanism. *J. Struct. Biol.* 132: 180–190. <http://dx.doi.org/10.1006/jsb.2000.4318>.
 215. Dodds DR, Gross RA. 2007. Chemicals from biomass. *Science* 318: 1250–1251. <http://dx.doi.org/10.1126/science.1146356>.
 216. Umezawa T. 2010. The cinnamate/monolignol pathway. *Phytochem. Rev.* 9:1–17. <http://dx.doi.org/10.1007/s11101-009-9155-3>.
 217. Bunzel M. 2010. Chemistry and occurrence of hydroxycinnamate oligomers. *Phytochem. Rev.* 9:47–64. <http://dx.doi.org/10.1007/s11101-009-9139-3>.
 218. Abokitse K, Wu M, Bergeron H, Grosse S, Lau PC. 2010. Thermostable feruloyl esterase for the bioproduction of ferulic acid from triticale bran. *Appl. Microbiol. Biotechnol.* 87:195–203. <http://dx.doi.org/10.1007/s00253-010-2441-6>.
 219. Koseki T, Hori A, Seki S, Murayama T, Shiono Y. 2009. Characterization of two distinct feruloyl esterases, AoFaeB and AoFaeC, from *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* 83:689–696. <http://dx.doi.org/10.1007/s00253-009-1913-z>.
 220. McAuley KE, Svendsen A, Patkar SA, Wilson KS. 2004. Structure of a feruloyl esterase from *Aspergillus niger*. *Acta Crystallogr. D Biol. Crystallogr.* 60:878–887. <http://dx.doi.org/10.1107/S0907444904004937>.
 221. Hermoso JA, Sanz-Aparicio J, Molina R, Juge N, Gonzalez R, Faulds CB. 2004. The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutive functional convergence in feruloyl esterase family. *J. Mol. Biol.* 338:495–506. <http://dx.doi.org/10.1016/j.jmb.2004.03.003>.
 222. Kroon PA, Williamson G, Fish NM, Archer DB, Belshaw NJ. 2000. A modular esterase from *Penicillium funiculosum* which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module. *Eur. J. Biochem.* 267:6740–6752. <http://dx.doi.org/10.1046/j.1432-1033.2000.01742.x>.
 223. Williamson G, Faulds CB, Kroon PA. 1998. Specificity of ferulic acid (feruloyl) esterases. *Biochem. Soc. Trans.* 26:205–209.
 224. Crepin VF, Faulds CB, Connerton IF. 2004. Functional classification of the microbial feruloyl esterases. *Appl. Microbiol. Biotechnol.* 63:647–652. <http://dx.doi.org/10.1007/s00253-003-1476-3>.
 225. Damasio AR, Braga CM, Brenelli LB, Citadini AP, Mandelli F, Cota J, de Almeida RF, Salvador VH, Paixao DA, Segato F, Mercadante AZ, de Oliveira Neto M, do Santos WD, Squina FM. 2013. Biomass-to-bio-products application of feruloyl esterase from *Aspergillus clavatus*. *Appl. Microbiol. Biotechnol.* 97:6759–6767. <http://dx.doi.org/10.1007/s00253-012-4548-4>.
 226. Grassick A, Murray PG, Thompson R, Collins CM, Byrnes L, Birrane G, Higgins TM, Tuohy MG. 2004. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. *Eur. J. Biochem.* 271:4495–4506. <http://dx.doi.org/10.1111/j.1432-1033.2004.04409.x>.
 227. Knott BC, Haddad Momeni M, Crowley MF, Mackenzie LF, Gotz AW, Sandgren M, Withers SG, Stahlberg J, Beckham GT. 2014. The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J. Am. Chem. Soc.* 136:321–329. <http://dx.doi.org/10.1021/ja410291u>.

Fernando Segato is an Assistant Professor in the Department of Biotechnology at the Engineering School of Lorena at the University of São Paulo, Brazil. He received his Ph.D. in Molecular Biology and Genetics of Microorganisms from the Ribeirão Preto School of Medicine at the University of São Paulo. His research focuses on prospection of hydrolytic and oxidative enzymes applied in lignocellulose depolymerization, exploration of thermophilic fungal genomes, and improvement of fungal cell factories.

André R. L. Damásio is a Young Research Scientist and works at the Brazilian Bioethanol Science and Technology Laboratory (CTBE) at the Brazilian Center for Research in Energy and Materials (CNPEM) in Campinas, Brazil. He received his Ph.D. in Biochemistry from the Ribeirão Preto School of Medicine at the University of São Paulo. His research focuses mainly on the improvement of *Aspergillus* strains for secretion of client proteins and the function of glycosylation in the folding and secretion of plant cell wall-degrading enzymes.

Rosy whole C. de Lucas is a Postdoctoral Fellow at Novozymes Brazil. She received her Ph.D. in Biochemistry from the Ribeirão Preto School of Medicine at the University of São Paulo. Currently, she is studying degradation of biomass cell walls by enzymes from microorganisms to improve second-generation (2G) ethanol.

Fabio M. Squina is a Senior Researcher at the Brazilian Bioethanol Science and Technology Laboratory (CTBE) at the Brazilian Center for Research in Energy and Materials (CNPEM) in Campinas, Brazil. He received his Ph.D. in Molecular Cell Biology from Ribeirão Preto School of Medicine at the University of São Paulo. His research interests combine molecular biology, “-omics” sciences, and high-throughput screening approaches, aiming to engineer enzymes and biotechnological routes for plant biomass conversion into bioproducts.

Rolf A. Prade is a Microbiology Professor in the Department of Microbiology and Molecular Genetics at Oklahoma State University in Stillwater, OK. He received his Ph.D. in Biochemistry from the Ribeirão Preto School of Medicine at the University of São Paulo. His research focuses on bioinformatics and molecular genetics of fungi, including carbon regulation and enzymatic pathways that metabolize plant cell wall polymers.